In compliance with the
Canadian Privacy Legislation
some supporting forms
may have been removed from
this dissertation.

While these forms may be included
in the document page count,
their removal does not represent
any loss of content from the dissertation.

University of Alberta

PULP MILL WASTEWATER TREATMENT PROCESS MODELING

by

Jun Wen

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of **Master of Science**

in

Engineering Management

Department of Mechanical Engineering

Edmonton, Alberta

Fall 2003

National Library
of Canada

Acquisitions and
Bibliographic Services

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale
du Canada

Acquisisitons et
services bibliographiques

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 0-612-87911-9
Our file Notre référence
ISBN: 0-612-87911-9

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou aturement reproduits sans son autorisation.

Canada

University of Alberta

Library Release Form

Name of Author: Jun Wen

Title of Thesis: Pulp Mill Wastewater Treatment Process Modeling

Degree: Master of Science

Year this Degree Granted: 2003

Permission is hereby granted to the University of Alberta Library to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only.

The author reserves all other publication and other rights in association with the copyright in the thesis, and except as hereinbefore provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatever without the author's prior written permission.

Date: September 24, 2003

University of Alberta

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled **Pulp Mill Wastewater Treatment Process Modeling** submitted by Jun Wen in partial fulfillment of the requirements for the degree of **Master of Science** in Engineering Management.

Dr. Peter C. Flynn

Date Sep. 24, 2003

ABSTRACT

An important goal of a pulp mill wastewater treatment plant is to efficiently monitor its process under different operating conditions in order to minimize the amount of effluent contaminants discharged into the receiving river. The objective of this research was to develop effective models to forecast the change of the amount of contaminants so that corrective actions can be taken in advance. Based on preprocessed field data, linear prediction error models were applied using selected variables. The shapes of oxygen uptake rate (measured by a respirometer) are quite sensitive to the biomass's activity in the activated sludge treatment process. A new variable was developed from the shapes' change. Adding the new variable, the models' prediction performances were improved.

To my family and friends

Acknowledgements

I would like to thank my supervisors, Dr. Ted Heidrick and Dr.Witold Pedrycz , for their guidance and support throughout my thesis work. Especially, I want to express my sincere gratitude to Dr. George Sedgwick for his encouragement, guidance and devotion to my research.

I thank Dr. Peter Flynn and Dr. Jozef Szymanski for providing many useful suggestions that enhanced the quality of this work. I also thank John Woolley and Mary Govenlock for providing valuable technical support. My gratitude also goes to Zhengang Han, Ying Li, Yimin Zhou and Dr. Dongguang Li for their help in this work.

The financial support of Natural Sciences and Engineering Research Council of Canada and Alberta Research Council is fully appreciated.

TABLE OF CONTENT

1	**INTRODUCTION**	**1**
	1.1 Background of the research	2
	1.2 ANC WWT process description	3
	1.3 Basic Activated Sludge Treatment concepts	6
	1.4 Current WWT control method applied by ANC	9
	1.5 Objective of the research	18
	1.6 Organization of this thesis	19
2	**LITERATURE REVIEW**	**21**
	2.1 Review of WWT process modeling methodologies	21
	2.2 Evaluation of different modeling methodologies	27
	2.3 Conclusion and Recommendations	35
3	**EMPIRICAL MODELING THEORY AND METHODOLOGY**	**37**
	3.1 Theory of linear prediction model	38
	3.1.1 General linear model	40
	3.1.2 Prediction Error Model	45
	3.2 Comparative analysis and recommended approaches	55
4	**PREPROCESSING OF THE DATA**	**58**
	4.1 Preprocessing of data	58
	4.1.1 Treatment of different sampling time	59
	4.1.2 Treatment of missing data	61
	4.1.3 Analysis of the trends	62
	4.1.4 Means removal	65
	4.2 Cross-correlation between the input and output variables	69
	4.3 Summary and Conclusions	77

5 MODELING RESULTS **78**

5.1 Using the conventional process data to predict FG TSS 85

5.1.1 One-input (AB pH) case 85

5.1.2 Two-input (AB pH, Settleability) case 90

5.1.3 Three-input (AB pH, Settleability, SC Torque) 93

5.2 Using the Respirometer data to improve the prediction of FG TSS 95

5.2.1 General analysis of OUR curve 95

5.2.2 Feature extraction of OUR curves 97

5.2.3 One-input (OUR Shape change) case 101

5.2.4 Two-input (AB pH & OUR Shape change) case 102

5.2.5 Three-input (AB pH, Settleability & OUR Shape change) case 103

5.3 Conclusion of model performance and its application 107

5.3.1 Comparison of modeling performance 107

5.3.2 Application of the model 109

6 CONCLUSIONS AND RECOMMENDATIONS **113**

BIBLIOGRAPHY **115**

APPENDIX

1 Detailed description of WWT variables 120

2 Modeling of FG TSS by different data partitioning 132

3 Using data from the 1st day to 419th day for modeling 136

4 Multi-Layer Perceptron (MLP) modeling to predict FG TSS 141

A 4.1 Theory of MLP Feed-forward Networks model 141

A 4.2 MLP modeling applied to predict FG TSS 148

LIST OF TABLES

1.1 Summary of process variables 10

2.1 Summary of reviewed WWT modeling and control methodologies 33

3.1 Summary of modeling methods 56

4.1 List of data of AB Setteleability from day 1 to day 20 60
4.2 Correlation of inputs to FG TSS 71

5.0 Experiments to be conducted by using different inputs 84
5.1 Performance index / model parameters (one input: AB pH) 85
5.2 Performance index / model parameters (two inputs: AB pH, Settleability) 90
5.3 Performance index /model parameters (one input: OUR shape change) 101
5.4 Performance index /model parameters (two inputs: AB pH, OUR shape change) 102
5.5 Performance index /model parameters (three inputs: AB pH, Settleability, OUR shape change) 104
5.6 Summary of the performance of models 108

A2.1 Performance index / model parameters (one input: AB pH, longer testing set) 132
A2.2 Performance index / model parameters (two inputs: AB pH, Settleability, longer testing set) 133
A2.3 Performance index /model parameters (three inputs: AB pH, Settleability, OUR shape change, longer testing set) 134
A3.1 Performance index /model parameters (three inputs: AB pH, Settleability, OUR shape change, full data set) 136
A3.2 Performance index /model parameters (three inputs: AB pH, Settleability, OUR shape change, longer testing, full data set) 140

LIST OF FIGURES

1.1 ANC's Effluent Treatment Process Diagram 4
1.2 A photograph of the ANC's effluent treatment plant 6
1.3 Basic elements for biomass growth 7
1.4 The structure of a respirometer 15
1.5 Flow chart of this thesis 19

2.1 A completely mixed aerated tank 23
2.2 Typical process control of NH_3 by DO (Lindberg, 1997) 29

3.1 Typical process signal flow schematic 38
3.2 Input and output signals of a hairdryer 39
3.3 Plot of y and u of a hairdryer 42
3.4 Model vs. Measured output, real input for training 46
3.5 Real output vs. ARX model output, real input, residuals 48
3.6 ARX model residuals' auto-correlation and its cross-correlation to input 49
3.7 Real output vs. ARMAX model output, real input, residuals 50
3.8 OE model signal flow 51
3.9 Real output vs. OE model output, real input, residuals 52
3.10 Real output vs. BJ model output, real input, residuals 55

4.1 Converting the sample time of AB settleability from a half day to one day 61
4.2 Add in missing data of H_3PO_4 flow by linear interpolation 62
4.3 Cumulative mean of FG TSS 63
4.4 Cumulative mean of FG TSS from day 108 to day 419 64
4.5 FG TSS plots, before and after the mean is removed 66
4.6a OE Training of FG TSS before mean removed 67
4.6b OE Training of FG TSS after mean removed 68
4.7 Adding back the mean after model parameters are found 68

4.8 Comparison of AB pH and FG TSS 75
4.9 Scatter plot of AB pH and FG TSS, from day 108 to 419 76

5.1 The flowchart of linear prediction error model optimization 79
5.2 Using CC NH_3, Temperature, AB pH as inputs 82
5.3 Using NH_3 Flow, P Flow, Temperature as inputs 82
5.4 Using CC NH_3, WAS Flow, RAS Flow as inputs 83
5.5 OE Training set, one input: AB pH 87
5.6 OE Validation set, one input: AB pH 88
5.7 OE testing set, one input: AB pH 89
5.8 OE testing set residual analysis 89
5.9 OE Testing set, inputs: AB pH, Settleability 91
5.10 OE testing set residual autocorrelation, input 1's cross-correlation with the residuals 92
5.11 Input 2's cross-correlation with residuals 93
5.12 OE testing set, inputs: AB pH, Settleability, SC Torque 94
5.13 One cycle of OUR and Nutrients Change 96
5.14 OUR cycles and FG TSS 96
5.15 Ensemble Averaged OUR in June 1998 98
5.16 Ensemble Averaged OUR vs. OUR on 28th day 98
5.17 Averaged OUR vs. OUR on 28th day (all are normalized) 99
5.18 OUR shape change 100
5.19 OE testing set, one input: OUR shape change 102
5.20 OE testing set, two inputs: AB pH, OUR shape change 103
5.21 OE testing set, three inputs: AB pH, Settleability, OUR shape change 104
5.22 Three-input OE testing set residual autocorrelation, input 1's cross-correlation with the residuals 105
5.23 Three-input OE testing set, input 2's and input 3's cross-correlation with the residuals 106

5.24 The training set of OE model, three inputs: AB pH, Settleability, OUR shape change 107
5.25 Inputs needed to predict the FG TSS 111
5.26 Measured and Predicted value of FG TSS 111

A1.1 Primary Clarifier 120
A1.2 Equalize Basin 122
A1.3 Secondary Treatment: ASB, Secondary Clarifier 123
A1.4 Polishing Pond 124
A1.5 TSS to PC, Flow to PC, PC Torque, TSS out of PC 125
A1.6 LWW TSS, PC BOD5, NH_3 Flow, CC Temperature 126
A1.7 CC pH, CC NH_3, RAS Flow, CC NH_3 127
A1.8 Sludge Age, SVI, P-ortho, AB pH 128
A1.9 MLSS, AB Settleability, AB NH_3, H_3PO_4 Flow 129
A1.10 SC Torque, SC TSS, WAS Flow, SC BOD5 130
A1.11 FG TSS, FG pH, FG BOD5, FG NH_3 131

A2.1 OE Longer testing set, one input: AB pH 133
A2.2 OE Longer Testing set, two inputs: AB pH, Settleability 134
A2.3 OE longer testing set, three inputs: AB pH, Settleability, OUR shape change 135

A3.1 OE testing set, three inputs: AB pH, Settleability, OUR shape change, full data set 137
A3.2 The training set of OE model, three inputs, full data set 138
A3.3 Measured and Predicted value of FG TSS, full data set 139
A3.4 OE longer testing set, three inputs, full data set 140

A4.1 Neuron Node structure 142
A4.2 MLP approximator structure 143
A4.3 MLP structure for one input, one output 146

A4.4 RMSE vs. Number of hidden nodes 147
A4.5 Function vs. Model outputs 147
A4.6 Plot of FG TSS and Prediction of all periods 148
A4.7 MLP Training set, three inputs: AB pH, Settleability, OUR shape change, full data set 149
A4.8 MLP Testing set, three inputs: AB pH, Settleability, OUR shape change, full data set 150

LIST OF ABBREVIATIONS

AB	Aeration Basin
ANC	Alberta Newsprint Cooperation
ASB	Activated Sludge Basin
AST	Activated Sludge Treatment
BOD5	5-Day Biochemical Oxygen Demand
CC	Contact Chamber
DO	Dissolved Oxygen
FG	Final Grab, collecting data just before the effluent is discharged into river
MLP	Multi-Layer Perceptron
NH_3	Ammonia
OUR	Oxygen Uptake Rate
P	Phosphorous
PC	Primary Clarifier
RAS	Return Activated Sludge
SC	Secondary Clarifier
TSS	Total Suspended Solids
WWT	Waste Water Treatment

CHAPTER 1

INTRODUCTION

Pulping separates wood into discrete fibres by chemical and mechanical processes. The fibres are then matted together and combined with papermaking chemicals to make a variety of paper products. Chemicals are also used to bleach pulps to make the paper white (Health Canada, 1997). Wastewater (effluent) from pulp and paper mill processes contains many potentially hazardous chemicals that can profoundly affect the environment. The effects of this effluent depend on a number of factors: the pulping and bleaching process; the extent to which fibres and chemicals are retained, reused, or discharged into the environment; and the extent of effluent treatment by external waste management systems.

Increasing demands for the discharge of pure wastewater from the effluent of pulp mill wastewater treatment (WWT) plants into rivers call for processes ensuring the reliability of the purification treatment. The quality standards set for mill effluent by Alberta Environment are among the strictest in North America, particularly for materials that consume oxygen in rivers. The Alberta Newsprint Company (ANC) has been maintaining a high level of environmental integrity in its operations through the

combination of a benign pulping process, well-trained operators, and a highly effective multi-stage effluent treatment system.

This thesis focuses on the development and implementation of models to assist process monitoring to minimize the amount of pulp mill effluent contaminants discharged into the receiving river. This thesis investigates modeling and data analysis techniques that may be applicable to a specific case. The pulp mill WWT modeling is based on real industrial data. A respirometer of Alberta Research Council (ARC) that measures mainly the respiration rate (the rate of consumption of oxygen) of the biomass to reduce contaminants is investigated to improve the model's prediction performance. This chapter provides a brief introduction to the WWT system and activated sludge treatment (AST) process. The key objectives of the thesis are also stated.

1.1 Background of the research

This project originated from ANC's attempts to answer the question, "How can we monitor and control key process outputs so that corrective action can be taken before a problem arises?" Model-based prediction of key outputs of a WWT plant can provide an answer to this question, and designing a proper model to predict them is the first step. However, because of the high degree of complexity of the industrial WWT process and its heterogeneity and the incompleteness and imprecision of real WWT data, finding a suitable model is challenging. In addition to using conventional WWT

variables to model process outputs, respirometry variables were added as inputs in an attempt to improve the performance of the modeling. The data from the respirometer can provide "snap-shots" of current AST process behavior.

A set of AST process data was collected over the benchmarking period from June 1998 to August 1999 at the ANC. The data included all conventional measurements conducted by the pulp mill staff as well as new process variables based on respirometry analysis. Since the pulping process (i.e., the loading into the treatment system) did not change much after the benchmarking work was completed, this data set forms the basis of this project.

1.2 ANC WWT process description

The ANC WWT process analyzed in this thesis is shown in Figure 1.1 (based on the design manual of ANC's WWT plant). The corresponding measured variables are listed below for each process component. The designed effluent retention time in each process component is indicated in terms of days. Each variable is assigned a number shown at its left top corner. A detailed description of the operation of each system component and how each variable is obtained is presented in APPENDIX 1.

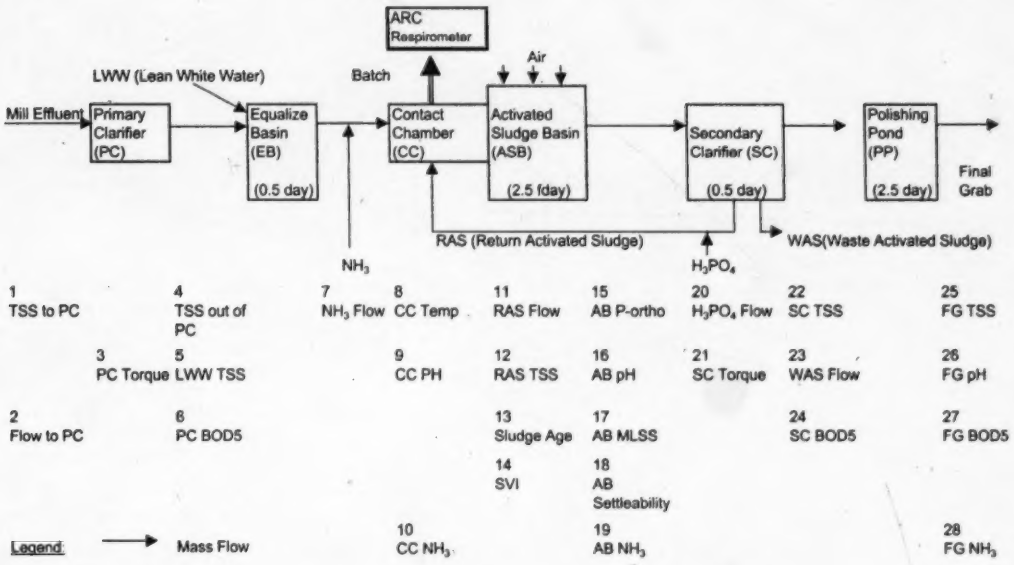


Figure 1.1 ANC's Effluent Treatment Process Diagram

As shown in Figure 1.1, after coarse screening that removes solids such as rags and wood, mill effluent containing organic substances, resin acids, pulps and possible toxic chemicals first enters the primary clarifier (PC) to remove some of the settleable solids. The presence of organic substances is indicated by the 5-day biochemical oxygen demand (BOD5) test. The BOD5 test measures the oxygen consumed by biomass as they decompose organics in a grab sample over a 5-day period. PC BOD5 is thus an indicator of the concentration of organics in water and is measured just after the PC (shown as variable #6 in Figure 1.1). The effluent then enters an equalize basin (EB) where it is retained for about 0.5 day to allow the basin's contents to be fully mixed and for the particles to settle. The retention times are based on the designed flow rates of the effluent and the volume of each process component to allow proper biochemical reaction and/or settlement of solids. At the same time, rather than creating

another WWT plant, another small portion (about 1/5 of the effluent flow) of lean white water (LWW, processing water from treating the pulp, containing little organic content as compared with the pulp mill effluent) also enters the EB. Effluent flowing out of the EB will be combined with the return activated sludge (RAS) in a contact chamber (CC). The RAS is a flow of sludge from the secondary clarifier (SC) and is continuously returned to maintain a high population of biomass. The effluent, RAS, and nutrients are very well mixed in the CC before entering into a clay-lined activated sludge basin (ASB). Two and a half days of hydraulic retention of the effluent provides sufficient time for the biomass to absorb the organics, oxygen and nutrients.

The activated sludge process is the key in removing the organic matter in the effluent. Aeration in the activated sludge basin is achieved by mixing the biomass with the wastewater and oxygen supplied by constantly blowing in air. The biomass grows by using oxygen and organic matter in the wastewater as food. As the biomass grows and is mixed by aeration, the individual biomass sticks together (flocculate) to form heavier clumps that thicken and settle in a quiescent tank (the secondary clarifier) in the form of settled sludge. Most of the settled sludge, so-called RAS, is returned to the ASB through the CC. As per the initial design, the effluent from the ASB flows by gravity to the secondary clarifier where it is retained for about 0.5 day to allow time for flocculent slurry to form for the proper settlement of the sludge. Then, to remove nutrient residuals and to handle plant upsets, the treated water enters a polishing pond (PP) where it is retained for 2.5 days to allow for the settling of solids carried over from the secondary clarifier or generated in the basin itself before being discharged

into the river. The effluent discharge is carefully monitored, and the key parameters such as BOD5, total suspended solids (TSS), and toxicity are measured in compliance with Alberta Environment requirements. The TSS is used to measure water quality as a concentration (weight of solids/volume of water; mg/L) of mineral and organic sediment. A photograph of the ANC's WWT plant is shown in Figure 1.2.

Figure 1.2 A photograph of the ANC's effluent treatment plant

1.3 Basic Activated Sludge Treatment concepts

As shown in Figure 1.3, oxygen, food, and nutrients are inputs for the healthy growth of the biomass (so-called bug, organisms or bacteria).

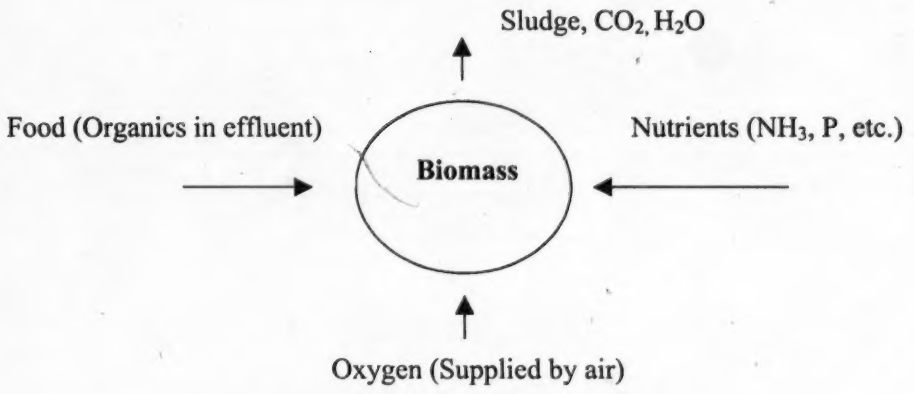


Figure 1.3 Basic elements for biomass growth

The objective of WWT process control is to monitor and control the oxygen and nutrient levels and the age and population of the biomass according to the effluent organic load to break down the organics effectively. The biomass must have a balanced diet to meet all of its nutritional needs. Otherwise, it will not thrive or, in some cases, may not even survive.

Industrial wastewater often lacks nitrogen and phosphorous, both of which are essential nutrients for the healthy growth of the biomass. Phosphorus plays a vital role in the energy transfer within the cells, while nitrogen is, in fact, a component in cell protein. Operating experience at numerous wastewater treatment facilities clearly shows the importance of having sufficient amounts of nitrogen and phosphorous available to the biomass. Nutrient availability and balance also may establish a "competitive advantage" for certain members of the microbial community, particularly when nutrients are limited. An insufficient supply of nutrient may lead to the growth

of nuisance biomass that interfere with the settling properties of activated sludge and may contribute to the production of troublesome scum, foam and excess slime. However, excessive nutrients will cause excessive nutrient residuals that are harmful to the receiving river. The biomass in the receiving river may thrive as a result of the excessive nutrients and thus deplete the oxygen in the river. Also, the nutrients are quite expensive. Therefore, maintaining a minimum nutrient and oxygen level according to organic load is very important. As shown in Figure 1.1, nitrogen is provided to WWT by using aqua ammonia (NH_3), and phosphorous is commonly provided in the form of phosphoric acid (H_3PO_4). Four factors are common to all activated sludge systems (Tchobanoglous & Burton, 1991):

(1) A flocculent slurry of biomass (mixed liquor [mixture of biomass and wastewater] suspended solids [MLSS]) is utilized to remove soluble and particulate organic matter from the influent waste stream.

(2) Quiescent settling for solids such as secondary clarifier is used to remove the MLSS from the process flow stream, producing an effluent low in suspended solids;

(3) Settled solids (sludge) are recycled as a concentrated slurry from the secondary clarifier back to the bioreactor in order to sustain a balanced biomass population;

(4) Excess solids are discharged to maintain the sludge retention time (SRT) at a desired value.

1.4 Current WWT control method applied by ANC

A key goal in wastewater treatment is to maintain effluent quality standards within certain constraints. Similar to control methods in other conventional pulp mills, ANC WWT control is realized through making comprehensive measurements and manipulating key variables based on known biochemical principles and experience. The control methodology is mostly case-based, using rules extracted from experience and biochemical principles to handle plant upsets case by case. Basically, traditional WWT control normally requires clear indications of process upsets at the output side of the plant before corrective action is taken. Certainly, by the time action is taken after a process upset, some pollution will have been discharged into the receiving water. For example, excess gray and thick slimy foam found on the surface around the weir of the ASB, would mean that nutrient deficiency had retarded the biomass growth. Therefore, the control action is to add more nutrients. However, the undergrowth of the biomass has already caused the improper treatment of organic effluent, which is harmful to the receiving water.

The ANC wastewater treatment facility interfaces with the mill's distributed control system (DCS). All key parameters are monitored in the central control room, as well as from a remote terminal in the treatment plant itself. The treatment facility, along with the water plant, is staffed with one operator per shift to conduct some of the routine operation testing requirements such as MLSS, sludge age calculations, sludge volume index (SVI), dissolved oxygen (DO), etc. Other testing (BOD5, TSS,

nutrients, etc.) is done in the main laboratory. Normally, the method of moving averages (7 and 28 days) is applied to smooth out the trends plot. The volume of information that is accumulated on a daily basis can overwhelm even the best process control engineer if he or she lacks a systematic way to manage the information. Ultimately, process control data need to be interpreted in a way that offers a clear picture of the plant's present status and provides a forecast of developing trends.

Table 1.1 Summary of process variables

Variable Category	Manipulative Variable	Final Grab (Discharge Test)	Monitoring variable in process
Organics		FG BOD5, FG BOD1	Effluent Flow to PC, LWW Flow before EQ, PC BOD5, LWW BOD5, SC BOD1, SC BOD5
Oxygen	Air flow rate	FG DO	OUR at CC, AB DO
Nutritions	NH_3 flow before CC H_3PO_4 flow at SC	FG NH_3	CC NH_3, AB NH_3 AB P-diss, P-ortho, P-total
Temperature		FG Temp	Temp to PC, Temp out of PC, CC Temp, AB Temp
pH	Lime added before PC	FG pH	pH to PC, pH to EQ, CC pH, AB pH
Sludge TSS Settleability	RAS flow at SC, WAS flow at SC, Sludge Age	 FG TSS	Flow PC Sludge, Sludge Bed Depth at SC, PC Torque, SC Torque TSS to PC, TSS out of PC, LWW TSS, TSS out of SC, RAS TSS, WAS TSS AB Settleability, SVI
MLSS			AB MLSS
Others		Conductivity, Colour	

Notes:
EQ: EQualization Basin
P-diss: Phosphorous -dissolved
PC: Primary Clarifier
Manipulative Variable: variables can be adjusted by operators directly

As show in Figure 1.1 and Table 1.1, the current WWT process control method applied by ANC is summarized as follows. The effluent is tested before discharge to

see if it meets government standards. Such testing also helps determine how efficiently the plant is operating. As shown in Table 1.1, typical effluent discharge tests include measurement of pH, DO, BOD and TSS. ANC also includes Temperature, NH_3 residual, conductivity and color in the discharge test. These are the control targets that the whole WWT process needs to be met. Also, the measurement of phosphorous residual at the weir of ASB is one of the target variables to be controlled.

The control target is achieved in stages by controlling the discharge test variables as well as the process monitoring variables through manipulating the manipulative variables (manually adjusted by operators to certain constant level). The organics, TSS, DO, temperature, pH and nutrients are the major variables to be controlled at all stages for the efficient activated sludge treatment. The following illustration is categorized according to the major controlled variables. Basically, all the control elements are to meet the needs of biomass in order to treat the effluent efficiently, and all the variables are interrelated with each other.

Temperature (Temp) has a direct effect on microbial metabolism. The rates of metabolic reactions, most of which are catalyzed by enzymes, tend to increase with increased in temperature, up to a limit. As the reaction rates increase, the rates of cell growth and reproduction also increase. Because growth rates and rates of removal of dissolved organics are directly related to temperature, the performance of biological treatment systems is affected by climate and seasonal temperature variations. Temp to PC, Temp out of PC, CC Temp, AB (Aeration Basin, also called ASB) Temp and FG

(Final Grab) Temp are monitored throughout the process. To maintain the temperature at the ASB weir, AB Temp is controlled at 29-35°C, and its limit is 27 and 37°C. Too high temperature will increase the turbidity with poor settling, and low temperature results in poor organics removal and can be observed by high oxygen concentration at AB weir. Spray cooler, fresh water/steam addition and heat exchangers are normally used to maintain the ASB temperature throughout the seasons.

Most biomass do well within a pH range of 6 to 8. Abnormal pH in biological treatment processes can result in a significant decrease in the rate of removal of organic compound from the environment. It can also select for the growth of unfavorable types of organism in the treatment process. It is very important to have the capability to control wastewater pH within an acceptable range before it is mixed with biomass in the ASB. Even short-term exposure (less than one minute) to extreme pH can cause significant microbial destruction. Lime (CaO) is added before the effluent entering a sump, which is before primary clarifier to control the pH of the effluent. Proper pH is maintained as constantly as possible in the whole WWT process.

The biomass converts soluble and colloidal mater (organics and suspended solids) into settleable sludge. For the organics, which is one of the objects that an activated sludge treatment plant needs to be gotten rid of, BOD5 is the key indicator of the concentration of organics in the effluent. The reduction of the BOD5 and the growth of the biomass are interrelated. They occur concurrently and represent the rate of removal of the carbonaceous material from the wastewaters or the treatment process.

Knowing the effluent and LWW flow rate, and the BOD5, the organic load to the WWT plant is the first thing to be monitored. At ANC, before and after the ASB, the PC BOD5 and SC BOD5 are monitored respectively. Other than BOD, TSS is another major criteria to measure the performance of WWT process, whether the colloidal mater and floc formed by biomass by converting organics are settled properly. TSS to PC, TSS out of PC, LWW TSS, TSS out of SC, RAS TSS, WAS TSS and FG TSS are monitored in the whole process. Before the treated effluent being discharged into the river, the FG BOD5 and FG TSS are obtained to make sure it meets government standard.

The biomass in the return activated sludge is essential for the proper removal of organics and suspended solids from the primary effluent. The biomass population is controlled mainly by the RAS flow. As shown in Figure 1.1, the RAS continuously returns biomass from the secondary clarifier to the ASB. RAS can be thought of as a solids balancing process in which sludge is recycled from the clarifier to the aerator.

Sludge age, also called sludge retention time or Mean Cell Residence Time (MCRT), which is controlled by the sludge wasting (removal of sludge biomass from the secondary system) rate, indicates the approximate number of days that the activated sludge remains in the system before being removed. Prolonged excessive sludge wasting will reduce sludge age by increasing the proportionate amount of newly developed floc in the system. On the contrary, improper low wasting rates will increase the number of days the sludge is retained in the system and will increase the

proportionate amount of older sludge. The goal of sludge wasting control is to develop a sludge quality that meets the performance requirements for a plant and stabilizes that sludge quality by maintaining the proper "sludge age" for the type, amount and quality of wastewater being treated.

All phases of microbial growth require sufficient oxygen to meet the demands of the process. Sufficient turbulence in the aeration tank must be maintained so that the biomass and wastewater are constantly mixed. Sufficient DO must be maintained throughout the aeration tank so that the entire floc particle is supplied with air. It is very important to routinely monitor the aeration basin for DO since it is one of the most critical elements in the activated sludge process. Too high of DO wastes energy and causes bug shift, floc breakdown, floating sludge (filamentous). Too low of DO has denitrification impact. Under anoxic conditions, nitrate can serve as a source of oxygen for a portion of the sludge mass. Nitrate respiration is called denitrification because nitrate is reduced to nitrogen gas. If denitrification lasts more than a day, low organics removal will come with high turbidity.

The biomass utilizes the oxygen to oxidize the organic food in the wastewater to provide organic carbon to synthesize the compounds necessary for their reproduction. The respirometer measures the rate of the utilization of oxygen for this growth process. As show in Figure 1.4, the respirometer, which is installed at ANC in 1997 (location shown in Figure 1.1) consists of a bioreactor containing a DO probe (to detect the dissolved oxygen residual).

The Tank Respirometer

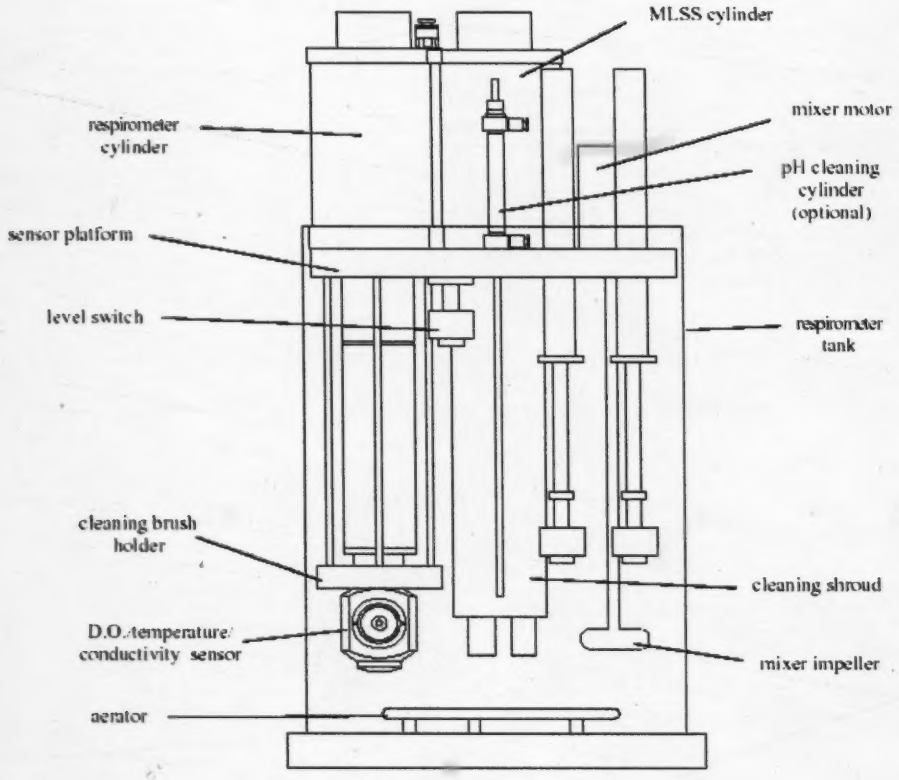


Figure 1.4 Schematic diagram of the respirometer used

The DO probe consists of a platinum cathode and a silver reference anode separated from the sample by a gas-permeable plastic membrane. A fixed voltage is applied to the platinum electrode. As oxygen diffuses through the membrane to the cathode, the oxygen is reduced. Accordingly, a current will flow that is proportional to the oxygen's diffusion rate, which is related to the concentration of dissolved oxygen in

the sample. This current is converted to a proportional voltage, which is amplified and read.

Shown in Figure 1.4, the respirometer has an integral cleaning mechanism, temperature thermistor, aerator, mixer, and level sensors. The tank is filled and emptied by using actuated valves. At the start of a run, the DO sensor is automatically wiped clean and exposed to air to enable an automatic recalibration of the DO sensor to take place. The bioreactor is then automatically filled with effluent sample directly from the activated sludge plant (as shown in Figure 1.1). The contents are then mixed. The aeration commences and continues until a user defined (higher) set-point (DO level) is reached. Aeration is stopped, but mixing and DO measurement continue. Aeration recommences when a lower set point is reached. Measurements are taken of the DO during the oxygen decay, and the oxygen uptake rates (OUR, the slope of the DO curve) are computed as the decay in the DO divided by the time lapse. The unit then empties, the DO sensor is cleaned and recalibrated, the respirometer refilled with sample, and the complete run is repeated.

Industrial wastewater often needs the addition of nutrients. Not only must the organic matter of wastewater now be eliminated as much as possible, but also the removal of nutrients needs to be considered. Ammonia is added at ANC WWT plant before the effluent entering contact chamber. It is important that monitoring for total nitrogen in the ASB effluent be performed regularly. The NH_3 is regularly monitored before and after activated sludge basin, and is carefully tested before the effluent is discharged

into river. Not enough NH_3 may cause unhealthy scam formation, and the overfeeding of NH_3 wastes money and sets up the potential for denitrification in the final clarifiers. Both nutrients should be fed proportional to the organic content. As a general rule, for each 100 mg/L BOD5, about 5 mg/L nitrogen and 1 mg/L phosphorous are needed. This is referred to as C:N:P ratio of 100:5:1. Normally, 0.5-1.0 mg/L phosphorous residual at AB weir is proper. If the phosphorous is too low, it will favor filaments leading to poor settling and high SVI. Certainly, we don't want to use too much expensive phosphorous.

In summary, all of the important control variables should be properly monitored and controlled to ensure the reliability of the purification treatment. Because of the high complexity and the heterogeneity of the WWT process, and the absence or imprecision of WWT data, finding a suitable model to aid the process control poses substantial problems. As a first step, the accurate prediction of key process outputs, or control variables, based on multivariate analysis is a central concern in order to satisfy the current discharge limitation.

1.5 Objective of the research

The overall objective of this research was to develop models to predict key process outputs of a pulp mill WWT plant. This allows corrective actions to be taken in advance to minimize the amount of contaminants discharged into the receiving river. In order to achieve the objective, the research was conducted in stages that each had the following objectives:

- To determine data analysis and modeling techniques that may be applicable to pulp mill WWT systems;
- To apply the selected techniques and models to existing industrial data to predict key process outputs;
- To evaluate the performance of the various models selected;
- To develop methods to most appropriately use the data measured by the respirometer.

The process followed to achieve these objectives is illustrated and discussed in the next section.

1.6 Organization of this thesis

This thesis is organized in the following manner presented in Figure 1.5.

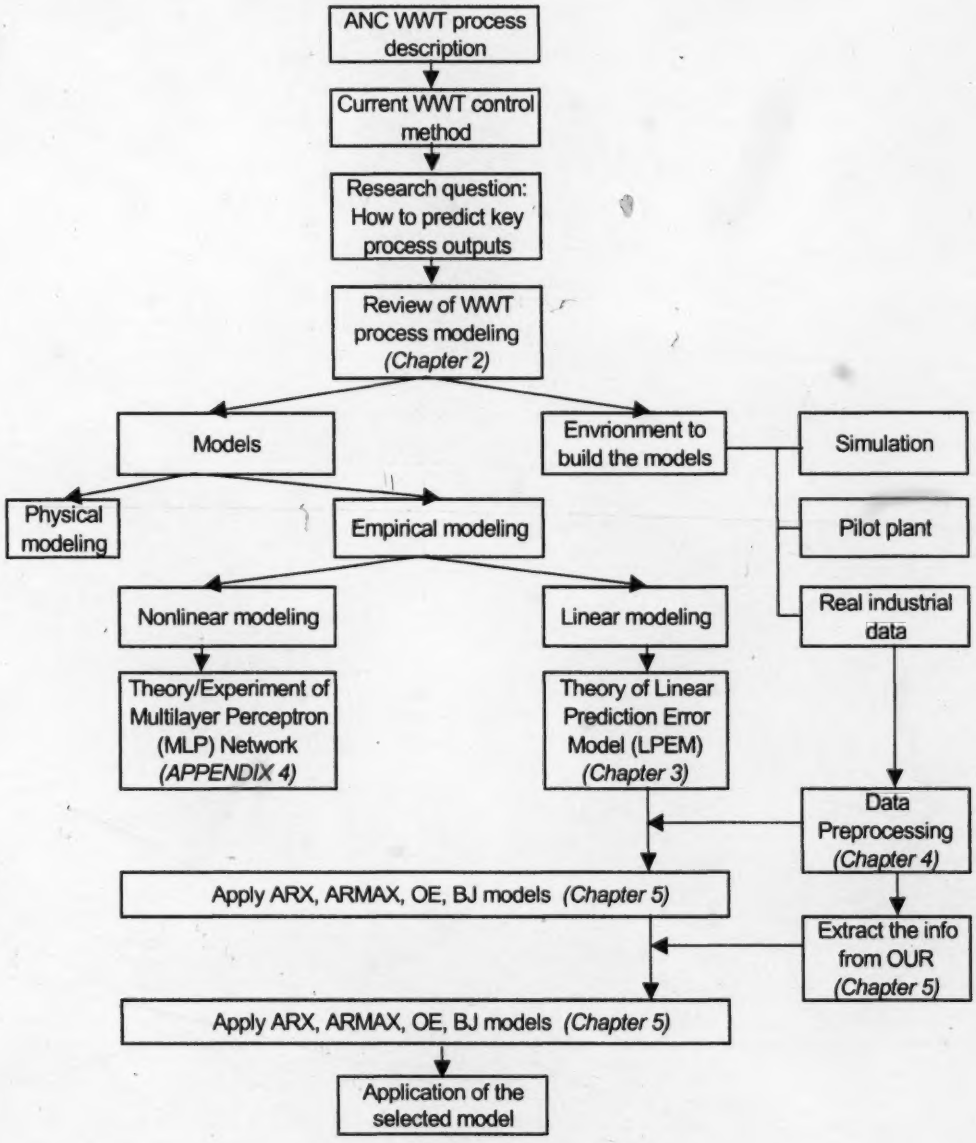

ANC WWT process description
Current WWT control method
Research question: How to predict key process outputs
Review of WWT process modeling (Chapter 2)
Models
Envrionment to build the models
Simulation
Physical modeling
Empirical modeling
Pilot plant
Nonlinear modeling
Linear modeling
Real industrial data
Theory/Experiment of Multilayer Perceptron (MLP) Network (APPENDIX 4)
Theory of Linear Prediction Error Model (LPEM) (Chapter 3)
Data Preprocessing (Chapter 4)
Apply ARX, ARMAX, OE, BJ models (Chapter 5)
Extract the info from OUR (Chapter 5)
Apply ARX, ARMAX, OE, BJ models (Chapter 5)
Application of the selected model

Figure 1.5 Flow chart of this thesis

Chapter 1 presents the background and basic concepts of the WWT process. Chapter 2 provides an extensive literature review, the results of which suggested the approach taken. Chapter 3 provides the theory of empirical models such as linear prediction error models that were subsequently used to analyze the data. Chapter 4 addresses the preprocessing of the data required before any modeling. Chapter 5 presents the results of the modeling. Chapter 6 draws the conclusions from this research and makes recommendations for future work.

CHAPTER 2

LITERATURE REVIEW

In this chapter, various WWT process modeling methodologies and categories of models are reviewed. The gaps in the research, research questions, and methodologies that could be further explored are analyzed. The usefulness of potential applications of WWT process modeling methodologies is presented.

2.1 Review of WWT process modeling methodologies

Since their inception by Arden and Lockett in 1914, the operation of activated sludge treatment plants has been widely investigated in the field of process control engineering. The typical modeling paradigms researched and applied for the identification and control of complex WWT process are physical modeling based on the physics of the phenomena and empirical modeling based directly on data from real plants. Simulations, pilot plants and industrial prototypes are commonly used to validate the models.

The increased public awareness of and more stringent regulations concerning plant emissions have considerably increased the requirements imposed on treatment plants compared with the requirements included in review compiled by Andrews (1974) some 30 years ago. Not only must the organic carbon pollution of wastewater now be eliminated, but also the removal of nutrients (e.g., nitrogen and phosphorous) needs to be considered. The numerous interactions that occur among the different unit processes and the fact that the biological potential is taken to its limits (e.g., by adding the minimum nutrients) make the WWT plants rather vulnerable to external disturbances or operation mistakes. Hence, the increased complexity because of this process integration demands the development of models, advanced instrumentation, and control systems (Vanrolleghem, 2000).

The activated sludge process is the key component in the WWT process to remove organic matter from the effluent. The following example, shown in Figure 2.1, uses a simplified mathematical model to illustrate the basic mechanisms of this process (Samuelsson, 2001).

In an activated sludge process, biodegradable matter is degraded by the biomass, which consumes oxygen. A general schematic of the process, simplified as a completely mixed bioreactor with one substrate, S (the material or substance such as organic matter on which an enzyme acts), and one biomass, X, is shown in Figure 2.1:

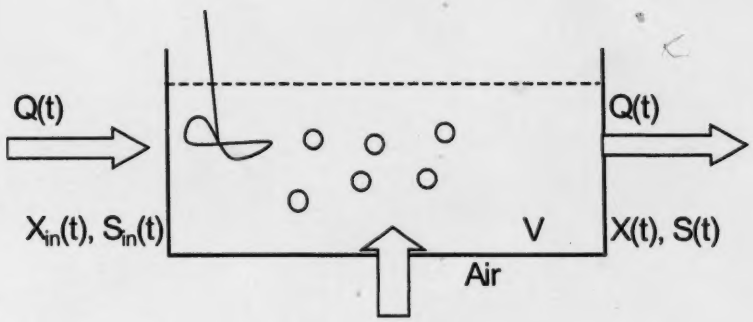


Figure 2.1 A completely mixed aerated tank

where Q(t) is the wastewater flow rate,

S(t) is the concentration of substrate,

S(t) is the concentration of substrate at the input of the aerated tank,

X(t) is the concentration of one kind of biomass, e.g., protozoa, at output

$X_{in}(t)$ is the concentration of one kind of biomass at the input of the aerated tank, and

V is the volume of the wastewater in the reactor.

A material balance is made for each substance of interest with the general form (Andrews, 1989),

Accumulation = Reaction + Input - Output

For the concentration of a process component, X, the simple mass balance yields the following dynamic model (Bastin and Dochain, 1990):

$$\frac{dX(t)}{dt} = R_x(t) + \frac{Q(t)}{V}(X_{in}(t) - X(t)) \qquad (2.1)$$

where $R_x(t)$ is the total reaction rate describing the biological processes.

Quite often, this reaction rate can be modeled by using the Monod function, also referred to as "Michaelis-Menten kinetics", which describes how the specific growth rate (the rate of increase in cell concentration per unit of cell concentration) of one component depends on the concentration of the different substrates. The Monod-function is a popular expression for simulating biological processes (e.g. biodegradation). For the simple one-substrate process shown in Figure 2.1, the specific growth rate is defined in the form,

$$\mu(S) = \mu_{max} \frac{S(t)}{K_S + S(t)} \qquad (2.2)$$

where $\mu(S)$ is the growth rate of biomass,

μ_{max} is the maximum growth rate, and

K_s is the half-saturation coefficient, that is the concentration that allows the biomass to grow at half the maximum growth rate.

From equation (2.2), it can be seen that the specific growth rate is zero if the substrate concentration, S(t), is zero and that the specific growth rate is approximately μ_{max} for large substrate concentrations. When the specific growth rate is given by equation (2.2), the total reaction rate for the biomass component X can thus be given by the assumption:

$$R_{X(t)} = \mu(S)X(t) \qquad (2.3)$$

Using the reaction rate in equation (2.3), the basic equation (2.1) will now yield the differential equation:

$$\frac{dX(t)}{dt} = \mu(S)X(t) + \frac{Q(t)}{V}(X_{in}(t) - X(t)) \qquad (2.4)$$

The above simple example only provides some basic understanding of bioreactor modeling. There are much more components could be incorporated into the building of more complicated models. A number of different models are used for the activated sludge process (ASP), but the most commonly used is the IAWQ (International Association on Water Quality) Activated Sludge Model No. 1 (ASM1) (see Henze et al. 1987). This model describes the removal of organic matter, process nitrification, and denitrification.

The models are primarily phenomenological because the rate expressions in them seek to depict the basic mechanisms involved. The simplifying assumptions normally made by the developers of the rate expressions and physical models are the following: (Tchobanoglous & Burton, 1991):

A. All of the biomass of a species are the same within a system. Actually, biomass differs physiologically because they are at different points in their lifecycles. Furthermore, little is known about the impact of their different states.

B. Stochastic phenomena are neglected, that is, any random differences among the cells of a given type are ignored. This is based on the assumption that the enormous number of cells present in most biochemical operations cancels out the effects of the random deviations.

C. Within a functional class, all biomass are treated as if they were the same species. This condition never exists in wastewater treatment systems.

D. The biomass are considered to be distributed evenly throughout the culture within a system.

In summary, microbial systems are extremely complex, and the models describing them can be very complicated. At the same time, the models are also empirical because the ultimate justification for the use of rate expressions is based on observation and experience rather than derivation from first principles. Both physical modeling and empirical modeling have certain advantages and limits. The choice should be made based upon the targets and resource constraints and after careful evaluation.

2.2 Evaluation of different modeling methodologies

It has been difficult to evaluate and compare the different control strategies of WWT plants, either in real application or by simulation because of the large variations in the influent, the complexity of the chemical and biochemical processes, and the great span of time constants (from a few minutes to several days) presented in the activated sludge process. An additional factor complicating the evaluation is the lack of standard evaluation criteria because regulations regarding effluent water quality are location specific. Each model found in the literature will be discussed in the section followed by the Table 2.1.

Rehnström (2000) used simulations to evaluate the control strategies applied to municipal wastewater treatment. A so-called "benchmark" was developed. The benchmark study utilized a simulated environment defining a plant layout, a process model, influent loads, test procedures and evaluation criteria. The objective of the different control laws was to control the nitrate and ammonia (NH_3) concentrations. The effluent water quality, assessed by the effluent quality index in the benchmark, was significantly improved when the control strategies developed in this study were applied.

The theme of Samuelsson's PhD thesis (2001) is the modeling, simulation and control of the activated sludge process as well as nitrogen removal in wastewater treatment plants based on ASM1. A model-based strategy to control the ammonium level by

manipulating the aeration volume (air flow rate) in the activated sludge process is developed. Simulations suggest that a combined feedforward-feedback strategy could give good results for controlling the aeration volume. Samuellson also suggests how to implement the control law in practice, i.e when the possible volume changes are discrete. Like Rehnström's study, Samuellson's thesis suggests that the next step could be to evaluate the suggested strategy in a pilot plant.

Lindberg's PhD thesis (1997) explores on-line methods for estimating the time-varying respiration rate and the nonlinear oxygen transfer function from measurements of the dissolved oxygen concentration and the airflow rate. The methods are successfully applied to real data. Methods for estimating the essential parameters characterizing the dissolved oxygen concentration dynamics in an activated sludge process are presented. The respiration rate is a key variable characterizing the DO process and the associated removal and degradation of biodegradable matter. A knowledge of the respiration rate is of interest in both process diagnosis (Olsson & Andrews 1978, Watts 1995) and process control, including dissolved oxygen control. Holmberg et al. (1989) suggested that the respiration rate is the only true indicator of the biologically degradable load. To find the respiration rate, one could, for example, do batch tests, use a respirometer, or apply an estimation algorithm (software sensor). Lindberg (1997) used the latter method. The estimation algorithm is based on measurements of the dissolved oxygen concentration and the airflow rate, both of which are used as inputs to a Kalman filter. Several estimation approaches have been suggested, e.g., by Holmberg et al. (1989) and Lukasse et al. (1996). Many of these

approaches have used simplified assumptions. For example, the respiration rate is assumed to be time invariant.

As shown in Figure 2.2, practical experiments in a pilot plant indicate that the ammonium concentration can be kept at a constant and low level, and at the same time, the energy consumption and the effluent nitrate concentration can be reduced.

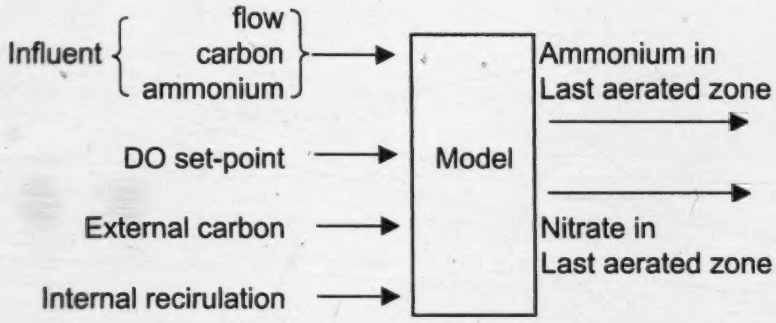


Figure 2.2 Typical process control of NH_3 by DO (Lindberg, 1997)

Figure 2.2 shows the input signals (the DO set-point, external carbon, and internal recirculation), measurable disturbances (influent flow, influent carbon and influent ammonium) and outputs (nitrate and ammonium in the last aerated zone) that the model is based on.

As an ultimate goal, plant-wide operation optimization is very complex. It is helpful to divide a plant's operations into several smaller sub-processes, which are easier to control and optimize. The global optimum operation of the plant may then not be reached, but getting close may be possible. Furthermore, to find the optimal operation

point for an activated sludge process, a cost function may be used. For example, by defining costs for effluent discharges and costs for running pumps, blowing air, dosing external carbon, etc., the total cost for a specific operating condition can be calculated. Vanrolleghem et al. (1996) have studied these matters.

Nejjari, et al. (1999) presented a nonlinear adaptive feedback-linearizing control designed for a biological wastewater treatment model. This adaptive control structure consists of a nonlinear controller combined with an estimation algorithm for the on-line estimation of biological states and uncertain parameters of the process. The nonlinear regulator is used to maintain the substrate and the dissolved oxygen concentration at a pre-specified level by acting on the dilution rate and the airflow rate. Simulation results with multiplicative measurement noises and parameter jumps are given to demonstrate the effectiveness and the robustness of both the estimation and control methods. However, many assumptions are made to simplify the model to formulate the differential equations.

An Artificial Neural Network (ANN) has the capability to model a process through extensive learning (Minsky & Papert, 1969; Rumelhart et al., 1986; Wasserman, 1989). For further details of ANN theory, refer to the Appendix 4. Researchers have shown that it can effectively handle the modeling problems associated with non-linear static or dynamic behaviors (Hoskins & Himmelblau, 1988; Bhat & McAvoy, 1990). Some key process variables of the WWT system are usually measured by laboratory tests, which typically take several days. A time-delay neural network (TDNN) (Zhu,

1998) modeling method has been proposed for predicting one of the WWT outputs, BOD5. Zhu performed a sensitivity analysis of a multi-layer perceptron (MLP) network (for further details, refer to Appendix 4) to reduce the input dimensionality of the model. The TDNN model was then further used to improve the performance of the original MLP network mode. Subsequently, an on-line prediction and model-updating strategy was proposed and implemented to better control of the WWT, a time-variant process. The modeling using industrial process data as inputs shows that the prediction accuracy can be improved by the on-line model updating. In future research, certain inputs, such as the respiration rate, which reflects the end result of several process variables, could be used to reduce the number of inputs. Too many inputs would increase both the amount of work required for data preprocessing and also the likely number of errors.

The use of on-line analyzers for continuous measurement of wastewater components is often accompanied by unpredictable breakdowns, leading to fault time in the measurements. A way to solve the problem of missing data is to use the known data to approximate missing ones. Hack and Kohne (1996) used an MLP with one hidden layer (for the theory, refer to Appendix 4) to allow for the approximate calculation of those process parameters that are temporarily not available.

Hennig (1999) presented a system that is suitable for the rapid analysis of a great number of samples. The very short measuring time and the low cost of measurements are the system's main advantages in comparison with biochemical standard analysis.

The inclusion of several methods allows for a useful interpretation of the complex interactions. Therefore, a multi-sensor system was developed, which can deal with processes that do not have to be completely understood. The evaluation model used in Hennig's measuring system is the MLP. However, the margin of prediction error varies between 5% and 25%, depending on the application. In some cases, this system can be used only for the estimation of orders of magnitude and in extreme cases, it allows for only yes or no answers regarding a certain state of the examined object. From a mathematical point of view, the output data become a representation of the input data. A simpler expression is that the network – when the training is completed – does pattern recognition on its inputs and tries to generate the desired output. Therefore, a successful simulation (model) requires a sufficient correlation between input and output data. Other than the MLP's configuration problems, insufficient input/output correlation could be the reason for an excessively high output error. Therefore, this method and its implementation of the MLP should be carefully compared with conventional analytical methods.

In summary, as shown in Table 2.1, for a focus on WWT process modeling, the empirical modeling basically includes linear modeling (Lindberg, 1998), nonlinear modeling (Nejjari, 1999), and ANN [Hack&Kohne(1996), Zhu(1998), Hennig(1999)]. Few examples of linear empirical modeling for predicting the WWT process outputs were found in the literature. In particular, the research about real industry process data that carry a great deal of noise and even incorrect measurements is quite limited. Lindberg's linear modeling (1997) was evaluated in a pilot plant study only.

Table 2.1 Summary of reviewed WWT modeling and control methodologies

	Objective	**Main Contribution**	**Method /Model**	**Limitation /Applicability to this thesis**
Rehnström (2000)	To control nitrate /NH_3 Concentration	Control strategies evaluation by benchmark study	Simulation in benchmark model	Research based on simulation only, has not been validated by pilot plant or industrial application yet.
Samuelsson (2001)	To control AST with nitrogen removal by manipulating aeration volume	To implement the control law when aeration volume changes are discrete	Simulation/ model based feedforward controller, exact linearization combined with some logical programming rules	
Lindberg (1997)	To keep effluent NH_3 and nitrates at low levels	Considers the respiration rate and oxygen transfer function in the Multivariable control design	Simulation/ Pilot plant/Linear Model	Based on municipal WWT, the model is validated by a pilot plant only. Not applicable to pulp mill WWT plants.

Nejjari (1999)	To maintain the substrate and DO concentrations by acting on the dilution & airflow rate	On-line estimation of the unavailable biological states and uncertain parameters	Mathematical nonlinear model/ Simulation /Nonlinear adaptive feedback control	Many assumptions are made to simplify the model
Zhu (1998)	To predict WWT result BOD5 by using industrial process data	Time-delay neural network and on-line model updating	MLP with on-line model updating / Sensitivity analysis to reduce input dimensions	Too many inputs would increase the work of data preprocessing and the likely number of errors
Hack & Kohne (1996)	To approximate those process parameters that are temporarily not available based on auxiliary parameters	Enables the detection of non-linear correlations based only on the measured values	MLP with one hidden layer based on industrial data	On-line updating could be used to enhance the practical application
Hennig (1999)	Fast analysis of a great number of samples for environmental monitoring, to predict NH_3 concentration	Revealed the importance of comparing the MLP with conventional analytical methods	MLP ANN modeling.	Further investigation on whether the high output error can be reduced by optimizing the MLP

2.3 Conclusion and Recommendations

Based on the above review and analysis, the following conclusion and recommendations are being made:

- The development of rigorous mathematical models may not be practical for complex biochemical processes if the models require a large number of differential equations with a significant number of unknown parameters.
- The literature reviewed focuses mainly on the modeling and control of BOD, NH_3, nitrate and DO concentrations. Forecasting of TSS (one of the key WWT process outputs) by linear prediction error model was not found.
- Watts's (1995) feature extraction of the oxygen uptake rate shape change is only qualitatively illustrated. To achieve efficient control of the WWT process, quantitive analysis of respirometry variables needs to be developed further. The in-depth and prompt information from the respirometer (for details, refer to Section 5.2) should support more accurate model based prediction control (MPC) of the WWT process.
- Simulations and pilot plants may be quite useful in validating models and setting up some basic control rules. However, a gap still exists between laboratory experiments and real industrial applications. Furthermore, since the location of various WWT plants are so different, a significant amount of resources are required to set up a pilot plant or a comprehensive simulation. Hennig (1999) demonstrates that an economical solution is to develop an empirical model directly from industrial data.

In conclusion, a practical strategy is to develop an empirical model directly from industrial data. In this project, empirical modeling was applied by treating the WWT plant as a black box whose inputs and outputs are known but whose inner workings are unknown. Traditional linear prediction error modeling was applied to the WWT plant data to develop a dynamic model to aid in the effective prediction of key process outputs.

CHAPTER 3

EMPIRICAL MODELING THEORY AND METHODOLOGY

Model Predictive Control (MPC) refers to a class of algorithms that compute a sequence of manipulated variable adjustments in order to optimize the future behavior of a process. Modeling is the foundation of MPC. A model is a mathematical abstraction of a real process. The set of equations comprising the model is at best an approximation of the true process. Hence, the model cannot incorporate all the features, both macroscopic and microscopic, of the real process. Engineers should seek a compromise between the cost of obtaining the models, that is, the time and effort required to obtain and verify them, and their performance. These considerations are related to the level of physical and chemical details in the models and the expected benefits to be derived from their use. A model's accuracy is necessarily intertwined in this compromise, and the ultimate use of the model influences how accurate it needs to be. Models can be considered in two different classifications, depending on how they are derived (Seborg et al., 1989):

a. Theoretical models developed by using the principles of chemistry, biochemistry and physics.

b. Empirical models obtained from a mathematical (statistical) analysis of process operating data.

The prohibitive expense and enormous time required to develop mathematical models of complex industrial processes from first principles often means that engineers must rely on system identification techniques to develop dynamic process models. A practical method for modeling a biochemical process is to develop an empirical model directly from experimental data. The most common approaches to system identification rely on linear models such as Auto Regressive with eXogeneous variables (ARX), Auto Regressive Moving Average with eXogeneous variables (ARMAX), Output Error (OE) and Box-Jenkins (BJ). Unfortunately, the majority of processes found in industries are non-linear, and in these cases, the performance of such models may be inadequate. As a result, efforts have also been directed towards the experiments one of the non-linear models, MLP modeling. In this thesis, empirical modeling was applied by treating the activated sludge treatment plant as a black box. The linear prediction error modeling gives quite useful results.

3.1 Theory of the linear prediction error model (Ljung, 1999, Soderstrom 1989)

Figure 3.1 illustrates a system with a scalar input signal u(t), a scalar output signal y(t), and a disturbance signal v(t).

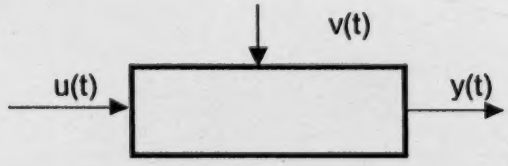


Figure 3.1 Typical process signal flow schematic

A system is said to be "time invariant" if its response to a certain input signal does not depend on time. It is said to be "linear" if its output response to a linear combination of inputs is the same as the linear combination of the output responses of the individual inputs. A system is said to be "causal" if the output at a certain time depends on the input up to that time only. A linear, time-invariant, causal system can be described by its impulse response $g(\tau)$ as follows:

$$y(t) = \int_0^\infty g(\tau)\, u(t-\tau)\, d\tau \qquad (3.1)$$

An example using data collected from a laboratory scale "hairdryer" is adopted to illustrate the theory. The process of the "hairdryer" works as follows: air is fanned through a tube and heated at the inlet. The input (u) is the voltage over the heating device, which is a mesh of resistor wires. The air temperature, the output (y), is measured by a thermocouple at the outlet.

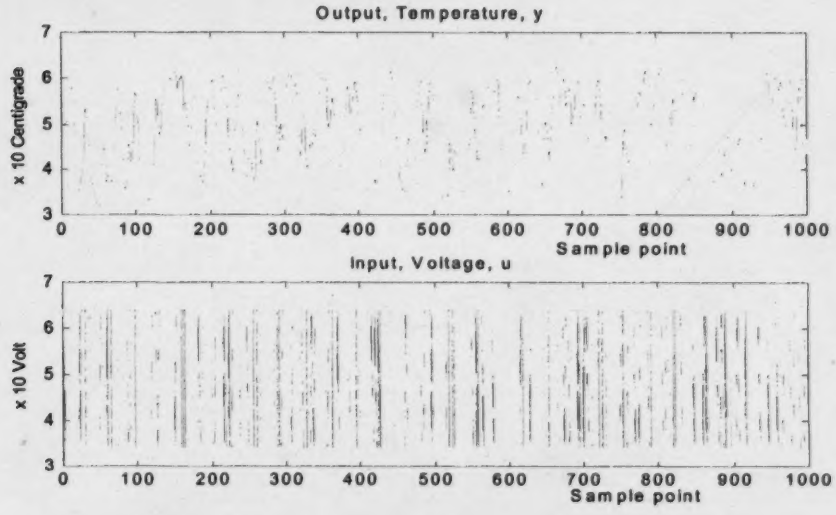


Figure 3.2 Input and output signals of a hairdryer

Shown in Figure 3.2 are y, the output, which contains 1000 measurements of temperature in the outlet air stream, and u, which contains 1000 input data points, consisting of the voltage applied to the heater. The input was generated as a binary random sequence that switches from one level to the other.

3.1.1 General linear model

In computerized control systems, we deal almost exclusively with observations of inputs and outputs in discrete time, since this is the typical data-acquisition mode. Refer to Figure 3.1, we thus assume y(t) to be observed at the sampling instants p=kT, k=1,2.... where T is the sampling interval. Normally, in computer control applications, the input signal u(t) is kept constant between the sampling instants. Thus:

$$y(p) = \sum_{k=1}^{\infty} g(k)u(p-k),\ p = T,2T... \qquad (3.2)$$

If we introduce the forward shift operator q by

$$q\,u(p) = u(p+1)$$

and the backward shift operator q^{-1}

$$q^{-1}\,u(p) = u(p-1)$$

we can then write (3.2) as

$$y(p) = \sum_{k=1}^{\infty} g(k)u(p-k) = \sum_{k=1}^{\infty} g(k)q^{-k}u(p) = \left[\sum_{k=1}^{\infty} g(k)q^{-k}\right]u(p) = G(q)u(p)$$

where we introduced the notation,

$$G(q) = \sum_{k=1}^{\infty} g(k)q^{-k}$$

which is called "the transfer function of the linear system" (3.2).

For ease of notation, we assume that T is one time unit and use "t" to express the sampling instants. According to the relationship (3.2), the output can be exactly calculated once the input is known as follows,

$$y(t) = \sum_{k=1}^{\infty} g(k)u(t-k), \quad t = 1,2... \tag{3.3}$$

However, in most cases, doing so is unrealistic. Signals beyond our control and affecting the system are always present. Within our linear framework, we assume that disturbances, signals beyond our control and affecting the system can be lumped into an additive term v(t):

$$y(t) = \sum_{k=1}^{\infty} g(k)u(t-k) + v(t), \quad t = 1,2...$$

Using the data from the "hairdryer" example, when t=1,2,3, the corresponding output, y, and input, u, are shown in Figure 3.3.

Assume u(0)=0,

when t=1,

$$y(1) = \sum_{k=1}^{\infty} g(k)u(1-k) + v(1)$$

$$y(1) = g(1)u(1-1) + v(1) = v(1)$$

when t=2,

$$y(2) = g(1)u(2-1) + g(2)u(2-2) + v(2) = g(1)u(1) + v(2)$$

and when t=3,

$$y(3) = g(1)u(2) + g(2)u(1) + v(3)$$

As shown in Figure 3.3, the linear modeling method uses y(3) to predict the measured value of $y_R(3)$ as closely as possible and assumes that the 3rd output y(3) is affected by u(1) and u(2).

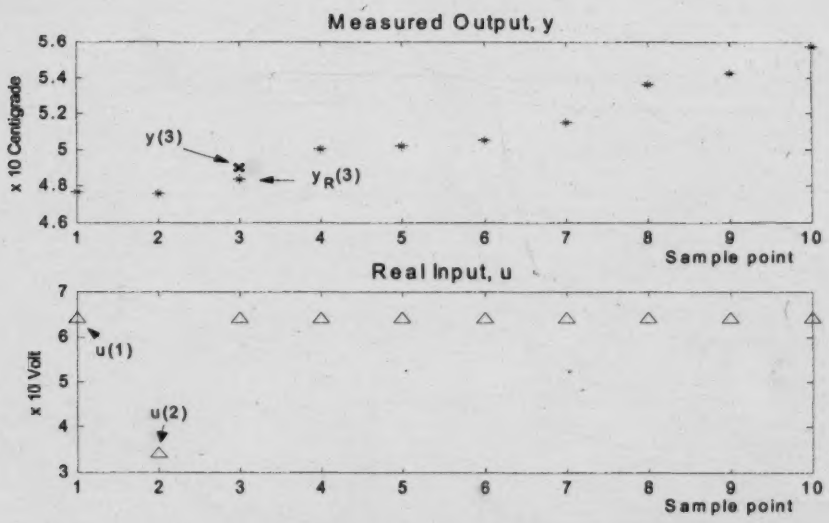


Figure 3.3 Plot of y and u of a hairdryer

The most characteristic feature of a disturbance is that its value is not known beforehand. Information about past disturbances (the past measured value minus the predicted value) could, however, be important for making qualified guesses about coming values. Thus, employing a probabilistic framework to describe future disturbances is natural.

Similar to (3.3), let v(t) be given as

$$v(t) = \sum_{k=1}^{\infty} h(k)e(t-k), t = 1,2,3,... \quad (3.4)$$

where {e(t)} is a sequence of independent random variables. Often, we specify only the second-order properties of the sequence {e(t)}, that is, the mean and the variances. The e(t) can be described as "a sequence of independent random variables with a zero mean value and variance λ". From this view of the system, v(t) can be considered as a disturbance signal, as shown in Figure 3.1. Although this description does not allow for completely general characterizations of all possible probabilistic disturbances in a process, it is versatile enough for most practical purposes.

The mean of v(t) can be calculated as

$$Ev(t) = \sum_{k=0}^{\infty} h(k)Ee(t-k), \text{ and assume } Ev(t) = 0,$$

and the covariance can be calculated as

$$Ev(t)v(t-\tau) = \lambda \sum_{k=0}^{\infty} h(k)h(k-\tau)$$

where h(r)=0 if r<0. We note that this covariance is independent of t and denote it as

$$R_v(\tau) = Ev(t)v(t-\tau) \quad (3.5)$$

the covariance function of the process v. This function, together with the mean, specifies the second-order properties of v. These are consequently defined by the sequence {h(k)} and the variance λ of e.

Similarly, with

$$H(q) = \sum_{k=0}^{\infty} h(k)q^{-k} \tag{3.6}$$

we can write,

$$v(t) = H(q)e(t) \tag{3.7}$$

Therefore, the basic description for a linear system with additive disturbance will be

$$y(t) = G(q)u(t) + H(q)e(t) \tag{3.8}$$

where G(q) is the transfer function of the system, and {e(t)} is a sequence of independent random variables with zero mean values and variance λ.

General linear models can be described symbolically by (3.8). The system identification problem states that the measured output y(t) is a sum of contributions from the measured input u(t) and contributions from the noise H(q)e(t). The symbol G(q) then denotes the dynamic properties of the system, that is, how the output is formed from the input. For linear systems, G(q) is called the "transfer function from input to output". The symbol H(q) refers to the noise properties and is called the "disturbance model". It describes how the disturbances at the output are formed from some standardized noise source e(t). As special cases of general linear models, four kinds of linear prediction error models are illustrated in the following sections. They have different features and can be applied in different situations.

3.1.2 Prediction Error Model

The simplest input-output relationship is obtained by describing it as a linear difference equation:

$$y\ (t) + a_1 y(t-1) + \ldots + a_{na} y(t-na) = b_1 u\ (t-1) + \ldots + b_{nb} u\ (t-nb) + e(t) \quad (3.9)$$

where e(t) is a sequence of independent random variables with zero mean values, also called white noise.

Since the white noise term e(t) enters here as a direct error in the difference equation, the model (3.9) is often called an equation error model (structure).

If we introduce $A(q) = 1 + a_1 q^{-1} + \cdots a_{na} q^{-na}$

and $B(q) = b_1 q^{-1} + \cdots b_{nb} q^{-nb}$

where na and nb are the order of the respective polynomials,

$a_1, \ldots, a_{na}, b_1, \ldots, b_{nb}$ are scalars, then,

$$A(q)y(t) = B(q)u(t) + e(t) \quad (3.10)$$

We will call the model (3.10) an "ARX model". To further illustrate the model, the example of the "hairdryer" is developed as follows. We select the first 300 values of the "hairdryer" for building prediction error models. First, we remove the constant levels and make the data zero-mean. The plots of the input and the model's vs. measured output are shown in Figure 3.4 (no unit is shown on the y axis after the values' means were removed). As shown, the model output is quite close to the measured output.

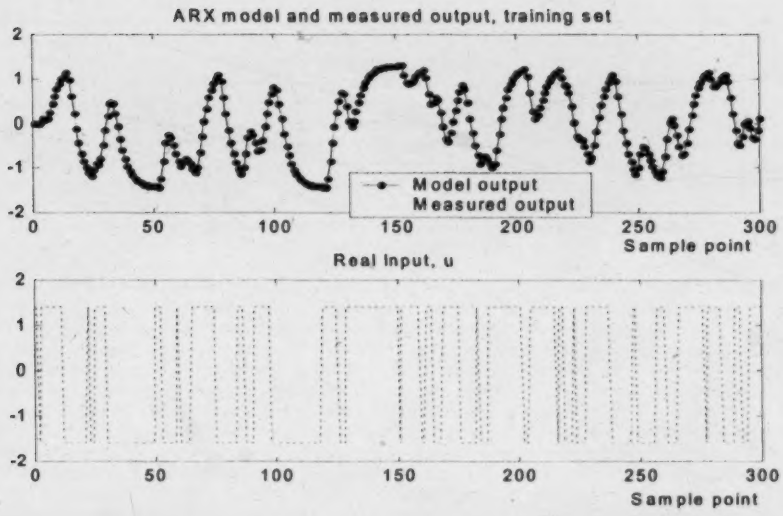


Figure 3.4 Model vs. Measured output, real input for training

The model with parameters na=2, nb=2 gives, relatively, the best performance among the lower order ARX models. The second order ARX model with delays of 3 sampling times may be computed in the form,

$$y(t) + a_1 y(t-1) + a_2 y(t-2) = b_1 u(t-3) + b_2 u(t-4) + e(t)$$

where y(t) is the model output,

u(t) is the real input, and

e(t) is the disturbance term in the model.

How good is this model? One way to find out is to simulate it and compare the model's output with the measured output. We then selected a portion of the original data that were not used as testing data to test the model's performance. We used samples 310 to 330 as testing data and simulated the output of the model with input u.

A comparison of the simulated output from the model with the measured output is shown in Figure 3.5. In Figure 3.5, particularly, when t=5,

$$y(5) + a_1 y(5-1) + a_2 y(5-2) = b_1 u(5-3) + b_2 u(5-4) + e(5)$$

$$y(5) = -a_1 y(4) - a_2 y(3) + b_1 u(2) + b_2 u(1) + e(5)$$

For a dynamic system with output lags of 3 sampling times to the input, the model output y(5) is assumed to be affected by y(4), y(3), u(1), u(2) and e(5). As Ljung (1999) notes, a dynamic system is the one that its current output value depends not only on the current external stimuli but also on their earlier values. For the physical mechanisms of a biochemical process, the dynamic model is quite suitable. For the example here, the output value at time t=5, say y(5), is built up from y(4), y(3), etc. In other words, y(5) is affected by y(4), y(3), etc.

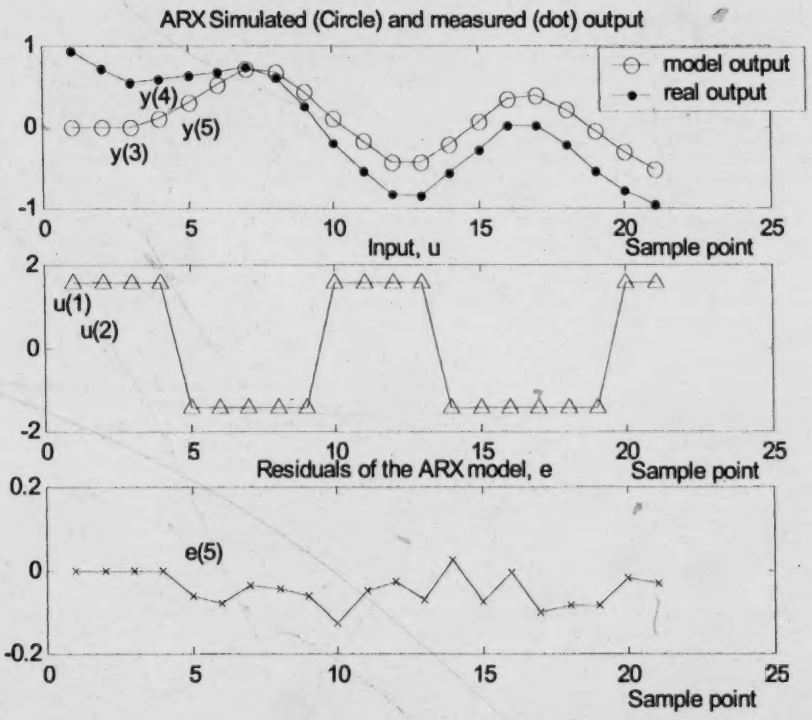


Figure 3.5 Real output vs. ARX model output; real input; residuals

As mentioned above in (3.9), e(t), the difference between measured and ARX model predicted value, is assumed to be white noise and are independent of the inputs. The following test is carried out to verify this. As shown in Figure 3.6, the residuals (measured – predicted values of the testing set) are uncorrelated among themselves. 99 % confidence limits (the dot line) for these values are also given. Since some of the correlation coefficients between input and residuals fall out of the confidence limits, we cannot treat the residual as white noise. Therefore, the ARX model is not feasible to be used for modeling here.

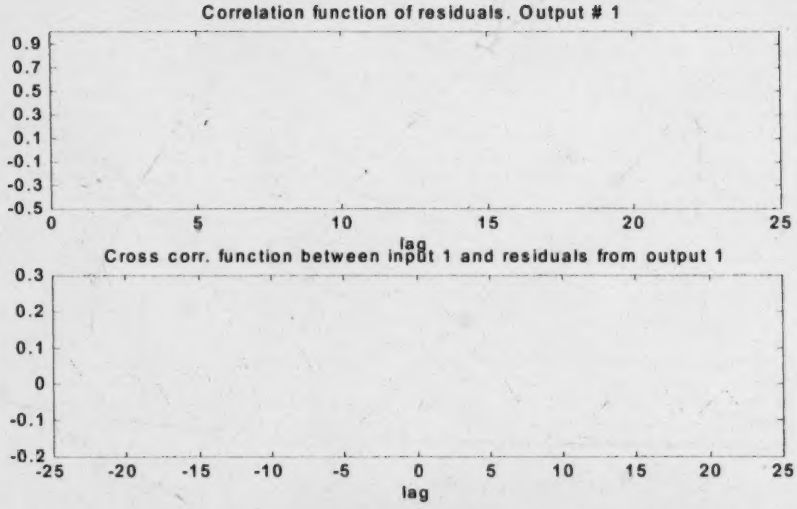


Figure 3.6 ARX model residuals' auto-correlation and its cross-correlation to input

ARMAX Model Structure

The basic disadvantage of the ARX model documented previously is the lack of adequate freedom in describing the properties of the disturbance term, e(t). We could add flexibility to it by describing the equation error as a moving average of white noise to characterize certain processes with more complex stochastic natures. Doing so gives the model,

$$y(t) + a_1 y(t-1) + \ldots + a_{na} y(t-na) = b_1 u(t-1) + \ldots + b_{nb} u(t-nb) + e(t) + c_1 e(t-1) + \ldots + c_{nc} e(t-nc) \quad (3.11)$$

With $C(q) = 1 + c_1 q^{-1} + \cdots c_{nc} q^{-nc}$, (3.11) can be rewritten as

$$A(q)y(t) = B(q)u(t) + C(q)e(t)$$

For example, when t=5, na=nb=nc=2, with a delay of 3 sampling times,

$$y(5) + a_1 y(4) + a_2 y(3) = b_1 u(2) + b_2 u(1) + e(5) + c_1 e(4) + c_2 e(3)$$

$$y(5) = -a_1 y(4) - a_2 y(3) + b_1 u(2) + b_2 u(1) + e(5) + c_1 e(4) + c_2 e(3)$$

As shown in Figure 3.7, the output y(5) is assumed to be affected by y(4), y(3) and u(1), u(2). As in the ARMAX model, $e(t)+c_1e(4)+c_2e(3)$ rather than e(t) itself is applied to model the disturbance.

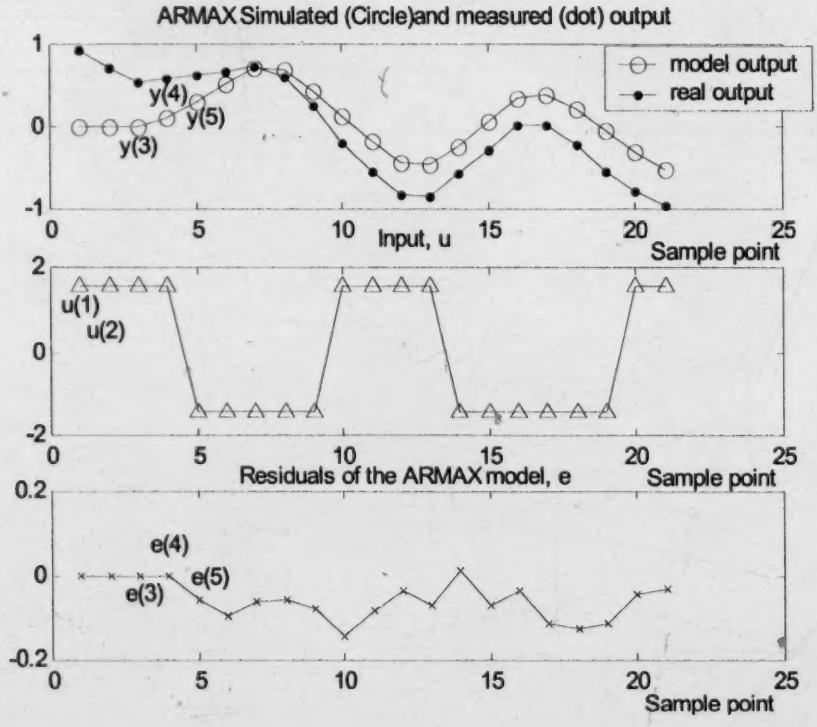


Figure 3.7 Real output vs. ARMAX model output; real input; residuals

OE(Output Error) Model Structure

Let $y(t) = w(t) + e(t)$

If we suppose that the relation between the input and the undisturbed output w can be written as a linear difference equation, and that the disturbances consist of white measurement noise, then we obtain the following description:

$$w(t) + f_1 w(t-1) + \ldots + f_{nf} w(t-nf) = b_1 u(t-1) + \ldots + b_{nb} u(t-nb)$$

With $F(q) = 1 + f_1 q^{-1} + \cdots f_{nf} q^{-nf}$

we get $F(q)w(t) = B(q)u(t)$, that is,

$$w(t) = \frac{B(q)}{F(q)} u(t)$$

and we can then write the model as

$$y(t) = \frac{B(q)}{F(q)} u(t) + e(t) \qquad (3.12)$$

We call (3.12) an "Output Error (OE) model" (structure).

The signal flow of the OE model is shown in Figure 3.8.

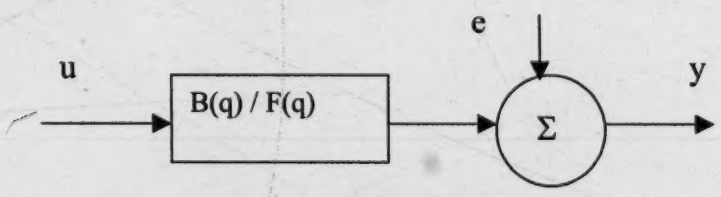


Figure 3.8 OE model signal flow

For example, when t=5, nb=nf=2, with a delay of 3 sampling times, we obtain

$$y(5) = \frac{b_1 u(2) + b_2 u(1)}{1 + f_1 q^{-1} + f_2 q^{-2}} + e(5)$$

As shown in Figure 3.9, the residuals of the OE model vary more significantly than those of the other three models because the OE model does not assume that the residuals are white noise.

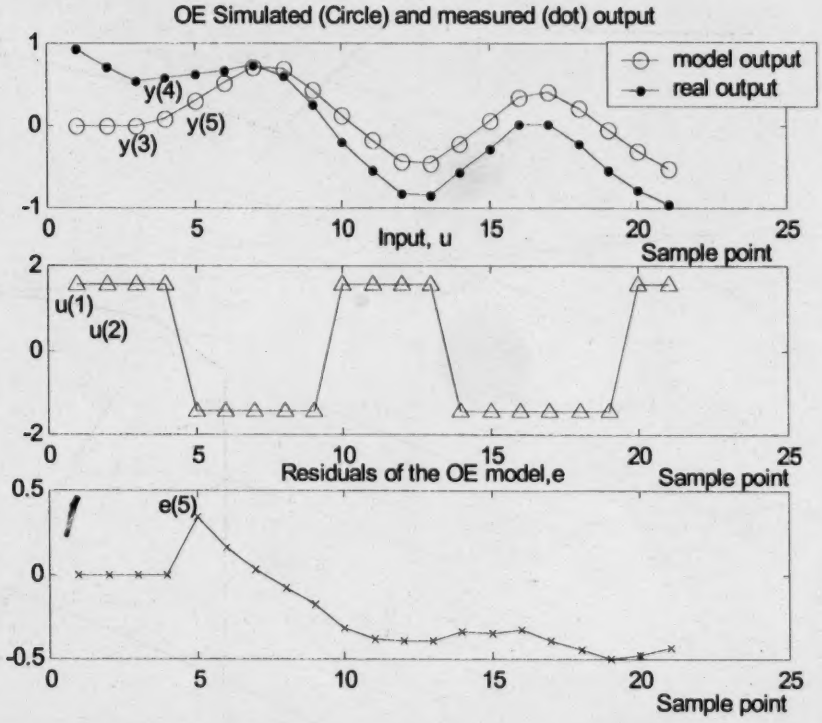


Figure 3.9 Real output vs. OE model output; real input; residuals

BJ (Box-Jenkins) Model Structure

A natural development of the output error model (3.12) is to further model the properties of the output error. That is,

$$y(t) = \frac{B(q)}{F(q)}u(t) + \frac{C(q)}{D(q)}e(t)$$

where, $D(q) = 1 + d_1 q^{-1} + \cdots d_{nd} q^{-nd}$

In a sense, this expression is the most natural finite-dimensional parameterization. Starting from equation (3.8), the transfer functions G and H are independently parameterized as rational functions.

In summary, for linear prediction error models, the ARX, ARMAX, OE and BJ model are developed from simple to comprehensive. This development is illustrated as follows.

The simplest prediction error model is described as an ARX model:

$$A(q)y(t) = B(q)u(t) + e(t)$$

Adding flexibility to describe the properties of the disturbance term by defining the equation error as a moving average of the white noise, the ARMAX model is developed as,

$$A(q)y(t) = B(q)u(t) + C(q)e(t)$$

that is, $y(t) = \frac{B(q)}{A(q)}u(t) + \frac{C(q)}{A(q)}e(t)$

However, this equation has the polynomial A(q) as a common factor in the denominators. From a physical point of view, parameterizing these transfer functions independently may seem more natural.

The BJ model realize this by defining $y(t) = \frac{B(q)}{F(q)}u(t) + \frac{C(q)}{D(q)}e(t)$

That is, for the undisturbed part of the output,

$$y(t) = \frac{B(q)}{F(q)}u(t)$$

$$y\ (t) + f_1 y(t-1) + \ldots + f_{nf} y(t-nf) = b_1 u\ (t-1) + \ldots + b_{nb} u\ (t-nb)$$

The equation error is defined as

$$y(t) = \frac{C(q)}{D(q)}e(t)$$

$$y\ (t) + d_1 y(t-1) + \ldots + d_{nd} y(t-nd) = c_1 e\ (t-1) + \ldots + c_{nc} e\ (t-nc)$$

As an example, when t=5, nb=nc=nd=nf=2, with a delay of 3 sampling times,

$$y(5) = \frac{b_1 u(2) + b_2 u(1)}{1 + f_1 q^{-1} + f_2 q^{-2}} + \frac{c_1 e(4) + c_2 e(3)}{1 + d_1 q^{-1} + d_2 q^{-2}}$$

The prediction performance and residual plots are illustrated in Figure 3.10.

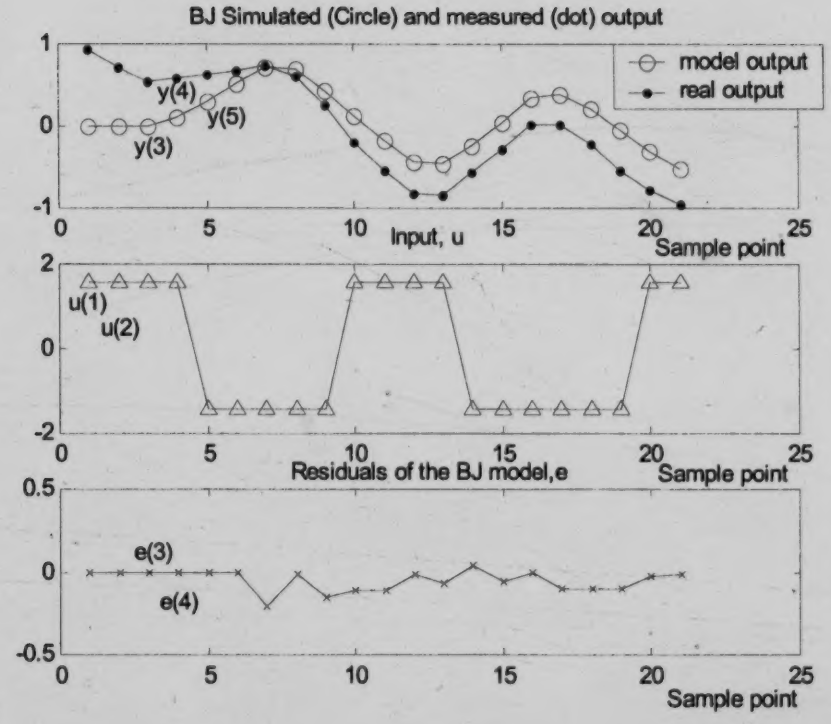


Figure 3.10 Real output vs. BJ model output, real input, residuals

3.2 Comparative analysis and recommended approaches

The features, advantages and disadvantages of the various models are summarized in Table 3.1.

Table 3.1 Summary of modeling methods

	Expression & Features	Advantage	Disadvantage
Model Type			
Physical Modeling	Using the principles of chemistry, biochemistry & physics. High level of physical/chemical detail	High Model Accuracy if complete	May need too many simplifying assumptions
Empirical Modeling	Obtained from a mathematical (statistical) analysis of process operating data. Treat the process as an opaque box.	Low cost of obtaining the model	Possible lower Model Accuracy
Linear Prediction Error Model (LPEM)	General Linear Model: $y(t) = G(q)u(t) + H(q)e(t)$	Can represent the dynamics of the process	Often does not adequately describe system dynamics at different operating points
ARX (Auto Regressive with eXogeneous variables)	$A(q)y(t) = B(q)u(t) + e(t)$ e(t): white noise		Lacks of freedom in describing properties of disturbance term Assume e(t) is white noise

ARMAX (Auto Regressive Moving Average with eXogeneous variables)	$y(t) = \frac{B(q)}{A(q)}u(t) + \frac{C(q)}{A(q)}e(t)$ Describing the equation error as a moving average of white noise: C(q)e(t)	More freedom to describe e(t)	Assumes input /disturbance affect the output in the same way: u(t),e(t) have a common denominator, A(q)
OE (Output Error)	$y(t) = \frac{B(q)}{F(q)}u(t) + e(t)$	Color noise, focus on modeling system dynamics	Lacks disturbance description factor: no polynomial to describe the relation of y(t),e(t)
BJ (Box-Jenkins)	$y(t) = \frac{B(q)}{F(q)}u(t) + \frac{C(q)}{D(q)}e(t)$	Most general LPEM	Most complex among LPEM, too many parameters may lead to less accuracy

As shown, there are basically two categories of modeling methods: physical modeling and empirical modeling. Since the mechanisms of biochemical processes are quite complicated, the empirical models were applied in the following sections based on the above analysis.

CHAPTER 4

PREPROCESSING OF THE DATA

A real industrial data set requires considerable preprocessing and careful selection of the modeling target in order to increase the efficiency of modeling. The reasons for this and the preprocessing methodologies are discussed below.

4.1 Preprocessing of data

The preprocessing of "raw" data should consider the following (Ljung, 1999):

A. Different sampling rates of various variables.

B. Missing data.

C. Slow Disturbances. Low-frequency disturbances, trends, drift, and periodic (seasonal) variations typically stem from external sources that we may or may not prefer to include in the modeling.

D. Removal of the means. Doing so allows the models to focus on modeling the dynamics of the process change around the set point in steady state.

Each aspect above is studied in the following sections.

4.1.1 Treatment of different sampling time

To treat the problem of different sampling rate, all data are converted to one sampling rate by the following methods. As shown in Figure 1.1, after the primary clarifier, the ANC's WWT plant takes on average 5.5 days to treat the mixed effluent, that is, 2.5 days to treat the effluent at the ASB, and 3 days to allow the effluent to flow through and to be treated in the secondary clarifier and polishing pond. If we use the variables obtained before the effluent enters the ASB to predict the ones after ASB, or if we use the variables just before the effluent enters the secondary clarifier to predict the output variables, which are measured after the effluent has left the polishing pond, a sampling rate of one day should be sufficient. However, if we choose a sampling rate of one week, the dynamics of the process cannot be described since most of the variables in the process change over the duration of a week. The designed retention time for the various components is in the order of days. The data based on weekly samples cannot be used in any predictive way. Daily data could be smoothed or averaged to produce a "weekly" number, but doing so would wash away the more frequent process fluctuations. For this reason, the nutrients' (NH_3 and P) residuals, which are sampled once per week, are not selected for analysis. It would make the already small data set even smaller if we choose the sampling rate of FG BOD5 (BOD5 measured just before the treated effluent to be discharged into the river), which is about three times a week. This is why the FG BOD5, a key process output, is not selected for analysis. For these reasons, only variables sampled daily, or more frequently than daily, were used to develop the model to predict one of the key process outputs, FG TSS.

Data sampled more frequently than daily are converted to those sampled once per day. For example, data sampled twice per day are converted to a sampling time of once per day by taking only one of the samples for each day. As shown by Table 4.1, the AB settleability (also called Settleability, measured at the output of ASB, Figure 1.1) is sampled by shift (sampling time =12 hours, one at 9AM, one at 9PM). The data sampled at 9AM are used since the other WWT data sampled daily are collected in the morning. Figure 4.1 illustrates the conversion of the sampling time based on Table 4.1.

Table 4.1 List of data of AB Setteleability from day 1 to day 20

Day	Date	AB Settleability
1	**11-Jun-1998 09:00:00**	**670**
2	11-Jun-1998 21:00:00	660
3	**12-Jun-1998 09:00:00**	**650**
4	12-Jun-1998 21:00:00	620
5	**13-Jun-1998 09:00:00**	**590**
6	13-Jun-1998 21:00:00	605
7	**14-Jun-1998 09:00:00**	**620**
8	14-Jun-1998 21:00:00	670
9	**15-Jun-1998 09:00:00**	**720**
10	15-Jun-1998 21:00:00	730
11	**16-Jun-1998 09:00:00**	**740**
12	16-Jun-1998 21:00:00	725
13	**17-Jun-1998 09:00:00**	**710**
14	17-Jun-1998 21:00:00	700
15	**18-Jun-1998 09:00:00**	**690**
16	18-Jun-1998 21:00:00	710
17	**19-Jun-1998 09:00:00**	**730**
18	19-Jun-1998 21:00:00	685
19	**20-Jun-1998 09:00:00**	**640**
20	20-Jun-1998 21:00:00	695

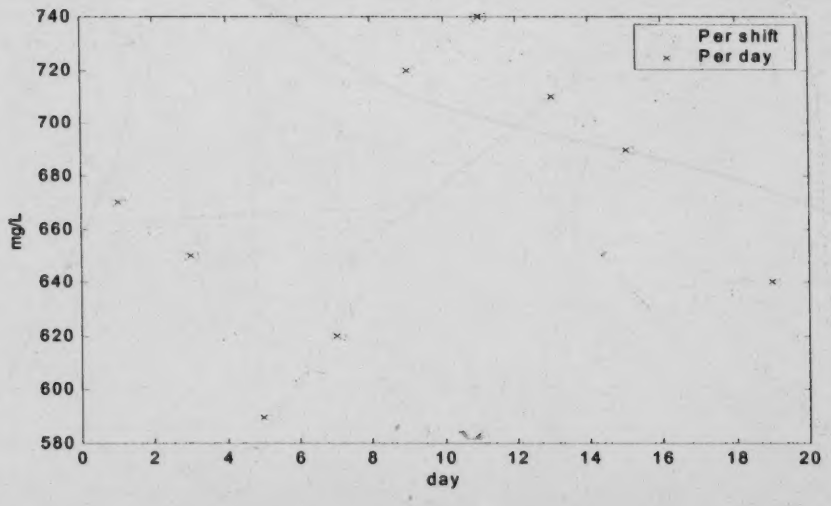


Figure 4.1 Converting the sample time of AB settleability from 0.5 day to 1 day

4.1.2 Treatment of missing data

For any variable missing one day of data, the method of linear interpolation is applied. For example, as shown in Figure 4.2, the H_3PO_4 flow rate is missing one data point on the 11^{th} of July, 1998. The average of the values for the 10^{th} and 12^{th} day is assumed to be the value of the 11^{th} day.

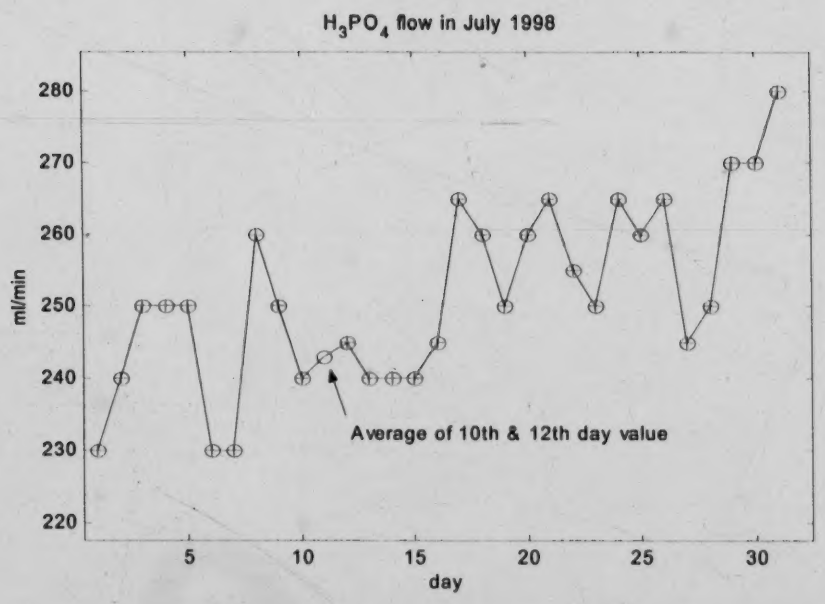


Figure 4.2 Add in missing data of H_3PO_4 flow by linear interpolation

4.1.3 Analysis of the trends

It is very important to remove trends and drifts when using methods employing a fixed noise model, such as output error methods. Otherwise, the discrepancy in levels will dominate the criterion of fit and the dynamic properties of the process will become secondary.

With an interval of one day, the cumulative mean plot of the key plant output, FG TSS, is shown in Figure 4.3. Each cumulative mean is calculated as follows. The 1st

datum is the average of day 1~1, the 2nd one is the average of day 1~2, the 3rd one is the average of day 1~3, ..., the 419th datum is the average of day 1~419.

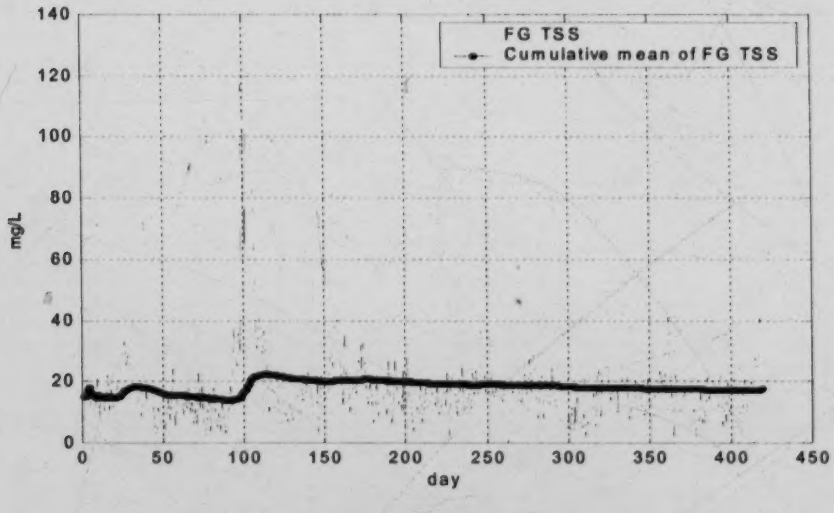


Figure 4.3 Cumulative mean of FG TSS

Relative to the magnitude of FG TSS, the variations in the cumulative mean are insignificant. The only major magnitude shift (about 5) happens around the 100th day because of a major plant upset. Since it is not the normal operating condition, we use the data from the 108th to the 419th day, or 302 points, for analysis. Similarly, the cumulative FG TSS mean plot is shown in Figure 4.4.

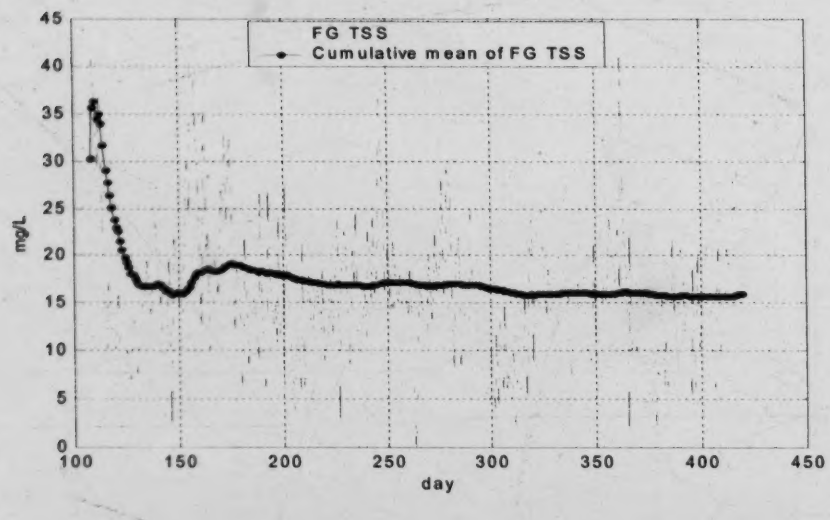


Figure 4.4 Cumulative mean of FG TSS from day 108 to day 419

The value of the cumulative mean converges to the value of 16 after about the 320th day. Therefore, except for the major plant upset, the whole process can be considered as a stationary process. The data from the 108th day to the 419th day are used in the following analysis. For reference purpose, the analysis using the data from the 1st day to the 419th day is reported in APPENDIX 3.

The likely physical mechanisms involved can further explain the stationarity of this process. This ANC wastewater treatment plant treats the effluent from a newsprint paper machine with a relatively consistent production rate. The grade of the paper is also relatively consistent. Thus, so is the content of the effluent. Also, the temperature of the effluent is observed to be consistent (at 33 ± 5°C) and is the major source (with a flow rate of about 10,000 L/min) that maintains the temperature in the ASB.

Furthermore, the temperature of the huge volume of the ASB, 37,500 m^3 was observed to remain at 32 ± 5°C all year round without obvious seasonal variation. Based on the above analysis, we can consider the data stationary and remove the data's means.

4.1.4 Means removal

The means of the variables are treated as follows. When the data have been collected from a physical plant, they are typically measured in physical units. The significant level difference in these raw input and output measurements will waste some parameters of the model in order to fit the levels. For example, if the input varies from 1 to 10 while the output varies from 20 to 100, a parameter in a model needs to describe the difference in terms of the range, or the so-called levels. By deducting the mean for both inputs and outputs, the model can save one parameter and focus on modeling the dynamics of the process. Otherwise, the discrepancy in average levels will then dominate the criterion of the fit, and the dynamic properties of the process itself will become secondary. Also, the nonlinear process can be modeled linearly only if the process runs around steady state. By removing the means, we can narrow the model on modeling the dynamic changes around steady state.

For the process with known physical mean level, offsets are easily dealt with: subtract the physical mean levels from the input and output sequences before the estimation. Since such values are not known in this case, the sample means are subtracted before analysis. The sample means is defined as,

$$\bar{y} = \frac{1}{N}\sum_{t=1}^{N} y(t) \ , \ \bar{u} = \frac{1}{N}\sum_{t=1}^{N} u(t)$$

and we then use $y(t) - \bar{y}$ and $u(t) - \bar{u}$ in the modeling.

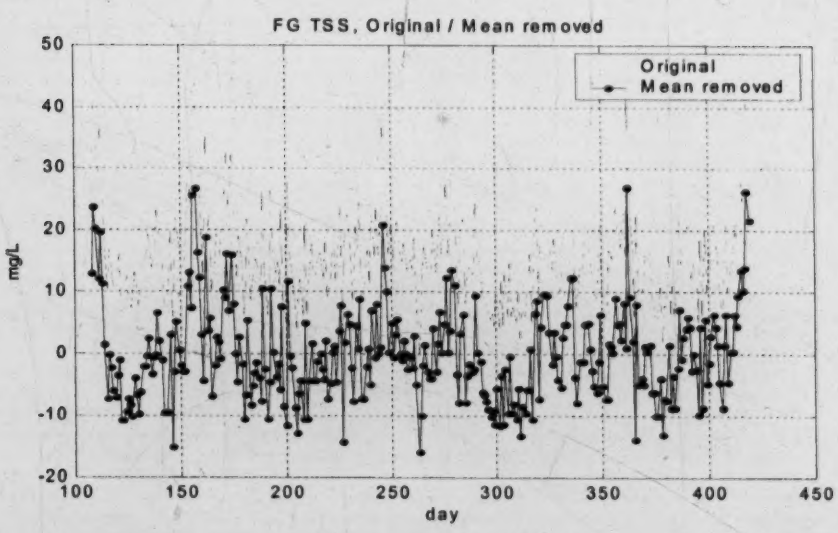


Figure 4.5 FG TSS plots, before and after the mean is removed

Figure 4.5 illustrates the FG TSS plot before and after the mean of FG TSS was removed. After the mean being removed, the FG TSS varies around the value of "0". For example, as shown in Figure 4.6a, without removing the mean, the OE model gives quite poor performance in the training set of FG TSS (from day 113 to day 310), the model output is quite close to a straight line. This is because that some of the OE model parameters are focused on the modeling of the level differences of the data. For example, around the 155th day, the change of FG TSS is 44 before mean is removed (Figure 4.6a), while the change of FG TSS is 27 after doing so (Figure 4.6b). The

bigger changes before mean removal will force some parameters of the model to fit them in terms of higher magnitude. That is, some of the model's parameters have to fit the model's level difference. As shown in Figure 4.6b, after removing the mean, the OE model performance is improved. The model output starts modeling the dynamic changes (no need to model the magnitude of the mean) of FG TSS. In this thesis, all means of the data had been removed before modeling.

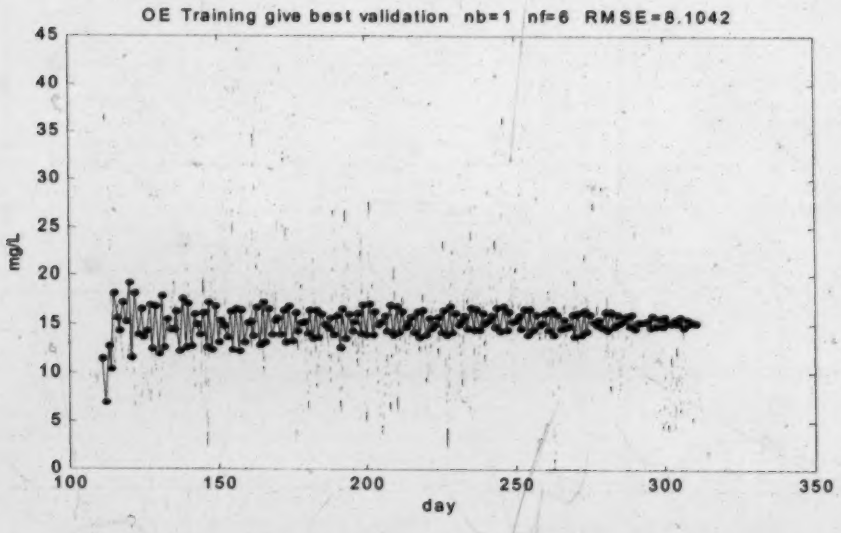


Figure 4.6a OE Training of FG TSS before mean removed

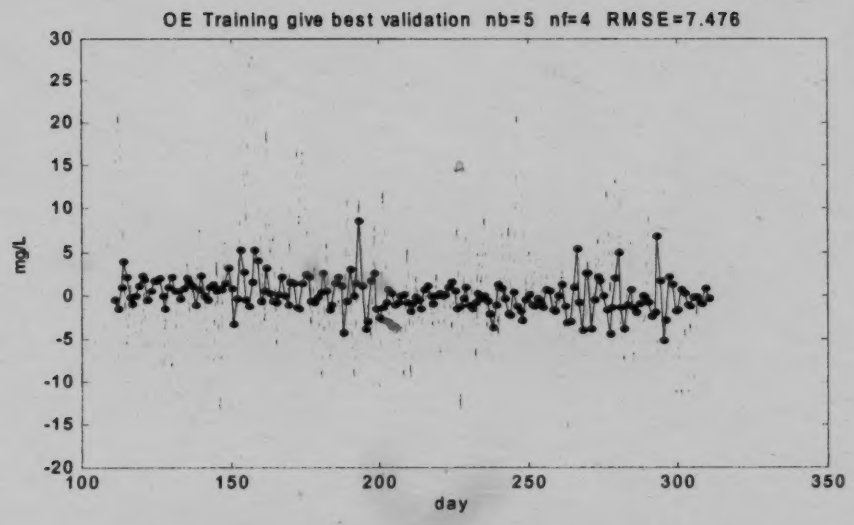


Figure 4.6b OE Training of FG TSS after mean removed

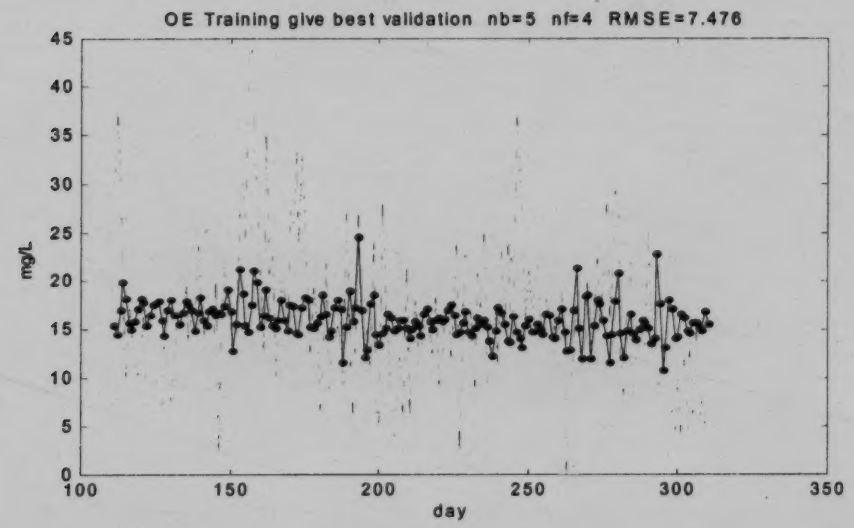


Figure 4.7 Adding back the mean after model parameters are found

After finding out the model parameters through model optimization, the mean was added back to the predicted model output in this thesis. That is, the means are added

back (original data's mean plus the model output by mean removed inputs) when plotting both of the model outputs and measured outputs (original data). It is essential to show the physical meaning of the outputs. For example, as shown in Figure 4.7, the mean of FG TSS is added back to the training set in order to show the physical meaning of the measured and predicted value of FG TSS.

4.2 Cross-correlation between the input and output variables

The optimization (to find out the best coefficients and orders of the linear prediction error model) of the linear model is conducted as follows:

Step 1: Find the correlations and delays between the inputs and outputs;

Step 2: Use the inputs that most correlated to the outputs to model the outputs of interest.

A correlation describes the strength of an association between variables so that the value of one variable can be predicted, to some extent, by the value of the other. A linear correlation is a special kind of association: a linear relation exists between the values of the variables. The coefficient of correlation can vary from a positive one (r=1, indicating a perfect positive relationship: when one variable goes up, another one goes up as well), to zero (indicating the absence of a relationship), to a negative one (r=-1, indicating a perfect negative relationship: when one variable goes up, another one goes down).

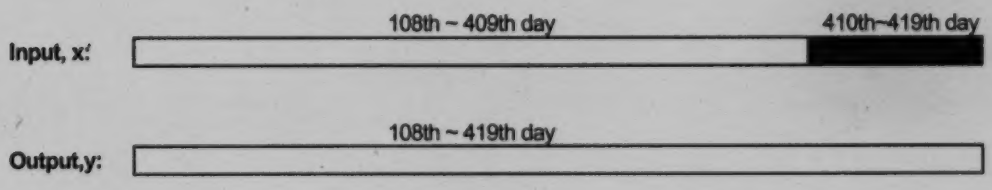
Cross-correlation is a measure of the liner dependency between two different data sets, computed by the sum of the cross products between the two data sets at different lag (i.e. time delay). The cross-correlation is a function of the lag, and is a standard method of estimating the degree to which two series are correlated. For two series x(i) and y(i) where i=1,2...N, consider the delays d=0,1,2,...N-1. The cross correlation series denoted by r(d) is defined as,

$$r(d) = \frac{\sum_{i=1}^{N}[(x(i)-\bar{x})*(y(i-d)-\bar{y})]}{\sqrt{\sum_{i=1}^{N}(x(i)-\bar{x})^2}\sqrt{\sum_{i=1}^{N}(y(i-d)-\bar{y})^2}} \qquad (4.3)$$

where $\bar{x}$ is mean of x(i), and $\bar{y}$ is mean of y(i), i = 1, 2, ..., N.

The denominator in the expression above serves to normalize the correlation coefficients such that $-1 \leq r(d) \leq 1$.

For the variables with a finite record length, since the data have different lags between them, the cross-correlation calculation should make certain that every point has a valid value to compute correlation based on part of the data. Using the data set of this thesis as an example, as shown below, both x (input) and y (output) have a record length of 312 data.



If y delays to x to a maximum of 10 sampling times, x is truncated so that it starts from the 108th datum and stops at the 409th (419-10 = 409) datum. The cross-correlation between x and y is calculated from a delay of one day to a delay of 10 days. Also, for the output y, refer to Figure 4.4, the cumulative mean of y reaches a stationary status after 320th day. Therefore, as a finite record length, data from 108th day to 409th day is long enough for calculating cross-correlation.

The cross-correlations of all the inputs to one output, FG TSS are calculated based on output lags to inputs from 0 to 10 days. All of the variables' correlations to FG TSS are calculated, with results ranging from 0.04 to 0.38. As shown in Table 4.2, AB pH (measured at the output of ASB, Figure 1.1), AB Settleability, SC Torque and FG pH are the highest correlated inputs (shown in bold print) to FG TSS.

Table 4.2 Correlation of inputs to FG TSS

	Delay (day)				
	1	2	3	4	5
AB_pH	-0.1822	-0.2012	-0.178	-0.1469	**-0.2084**
AB_Settleability	-0.1999	-0.2075	-0.2276	-0.2205	**-0.2416**
SC_Torque	-0.3141	**-0.3162**	-0.2997	-0.2759	-0.2616
FG_pH	**0.3787**	0.352	0.3134	0.2798	0.2161

Correlation, a statistical analysis, implies nothing about cause and effect, especially for multi-input and multi-output cases. If a correlation is found, the secondary question should be asked, "Does this correlation make any physical or biochemical

sense?" Since the FG pH is measured at the same point as FG TSS in the process, the FG pH cannot be used to predict FG TSS. The corresponding delay when the correlation value reaches the highest value cannot be blindly applied as the delay in calculating the model. For the multi-input multi-output cases, since the inputs and outputs interfere with each other, a correlation between only one input and one output normally cannot show the actual relation between them. Instead, the delay based on the retention time indicated by Figure 1.1, the designed process flow diagram, could be more meaningfully adopted. For example, the correlation coefficient of AB Settleability reaches the highest absolute value of about 0.24 when the delay is 5, and the second highest absolute value is about 0.23 when the delay is 3 days. Based on Figure 1.1, the delay of 3 days is consistent with the process's designed retention time. To further verify the above reasoning, both delays are used in the modeling, and the results are analyzed.

Although the cross-correlation provides some information about the plant dynamics, the magnitude of the correlation coefficient is not high. A question is thus raised, "Are these correlation coefficients statistically significant?" For example, the measure of AB pH is correlated with FG TSS with a coefficient of –0.178. The question is, "Does the value of the correlation coefficient indicate a dependency between the two variables in the population from which the sample was drawn? " (Johnson, 1939)

To answer the above question, we can perform a hypothesis test. Ronald A. Fisher (1890-1962) viewed hypothesis testing as a method for forming an opinion as to

whether or not a specific value of a parameter is likely to be the true value. Using Fisher's method, we can determine whether the correlation result could occur by chance (Byrkit, 1987). A null hypothesis is often the reverse of what we actually believe. That is, if we can conclude the hypothesis of no correlation is unlikely to be true, then we can say that the alternative hypothesis, that there is a correlation, is likely true. It is conducted by following three steps.

1. Specify the null hypothesis, that is, no linear correlation exists between the two variables, $\rho = 0$ (ρ, the linear correlation coefficient for the population), then specify the alternative hypothesis, $\rho \neq 0$, that there is a non-zero correlation.

2. Choose a significance level (also called the p-value; the reliability of a relation between variables observed in the sample can be quantitatively estimated and represented by using a standard measure). The p-value represents the probability of an error being involved in accepting our observed result as valid, that is, as "representative of the population." For example, a p-value of .05 indicates that a 5% probability that the relation between the variables found in the sample is by chance. We collect samples from populations and use the sample information to test hypotheses concerning the population as a whole. A 0.01 level of significance is used in the following analysis. Designate the test statistic as the t-distribution since we do not a priori know whether the value of the correlation, r, calculated from the sample, is normally distributed. The t-distribution approaches the normal distribution as the sample size increases.

3. Compute the p-value, the probability of obtaining a difference between the values specified by the null hypothesis is as large or larger than the difference obtained in the experiment. Apply the general formula for testing the hypothesis. Based on the correlation calculation, r = -0.178, since n=302, therefore,

$$t = \frac{r - \rho}{\sqrt{(1 - r^2)/(n - 2)}} = \frac{-0.178 - 0}{\sqrt{(1 - 0.178^2)/(302 - 2)}} = -3.65$$

The degree of freedom (df) is df = 302-2 = 300. Based on the Student's t-Distribution table (Byrkit, 1987), the absolute value of t (| t | = 3.65) for a two-tailed test (having the rejection area divided into equal portions in both tails of the distribution) is larger than the critical value for significance level of 0.01 (t value corresponding to 0.01 is 2.576 when df = 300). Therefore, the correlation is significant. Since the correlation is significant, the null hypothesis is rejected. It is concluded that the correlation between AB pH and FG TSS is significant. Similarly, the correlation between AB Settleability and FG TSS has a t value of –4.72, and the correlation between SC Torque and FG TSS has a t value of –6.34. Both correlations are statistically significant.

As shown in Figure 1.1, FG TSS is measured after polishing pond just before the treated effluent being discharged into the receiving river. Based on correlation analysis, the AB pH is correlated to FG TSS and is measured before the effluent enters the secondary clarifier. It is 3 days ahead of FG TSS, based on the hydraulic design of the WWT plant. As shown in Figure 4.8, the correlation can initially be inferred from the plots of input, AB pH and output, FG TSS from June 11, 1998 to August 03, 1999. A major negative spike of AB pH happens around 108th day and a strong positive

spike of TSS (highest magnitude=134.5 mg/L) happens around the 110th day because of a plant shutdown. In terms of its physical meaning, a sharp drop in the pH will not favor the growth of the biomass and thus will cause poor settlement and an increase in FG TSS value.

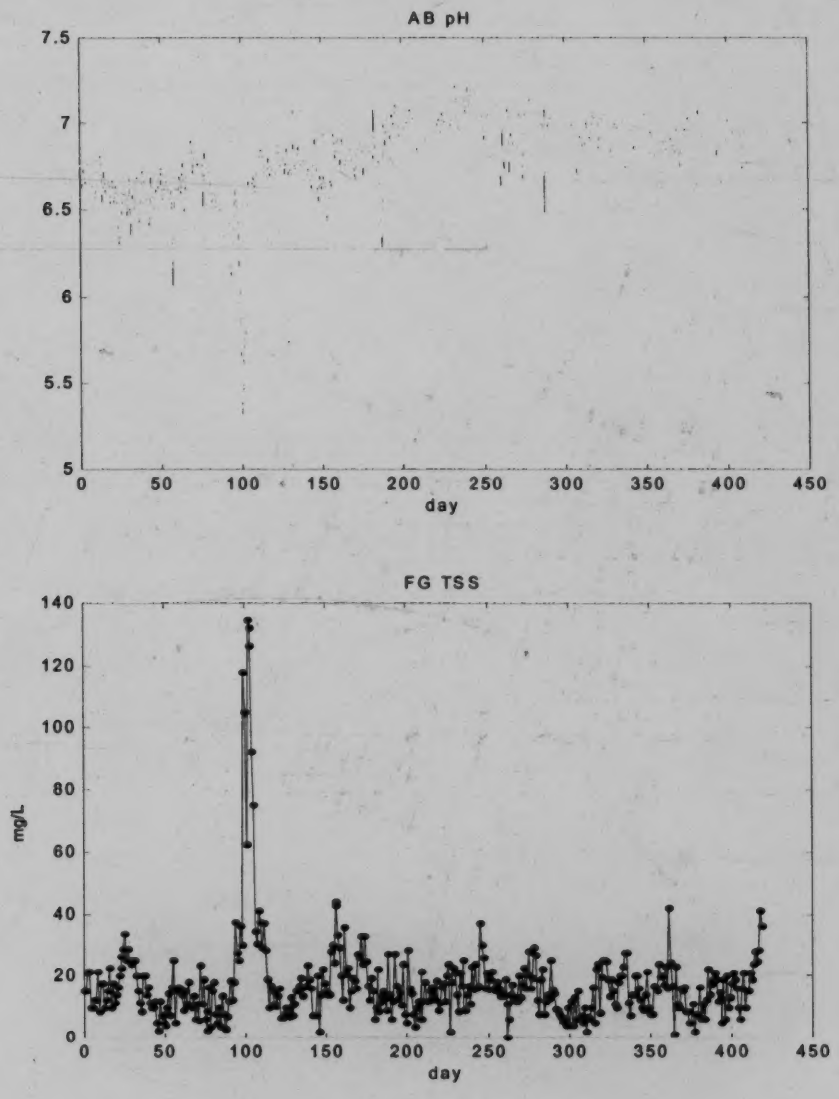


Figure 4.8 Comparison of AB pH and FG TSS

Figure 4.9 shows the scatter plot of AB pH and FG TSS. This figure indicates a weak linear correlation between them. Also it does not indicate that they have any other relation (e.g., if the plots form a curve, it shows a nonlinear relation). The maximum of their correlation is presented in Table 4.2.

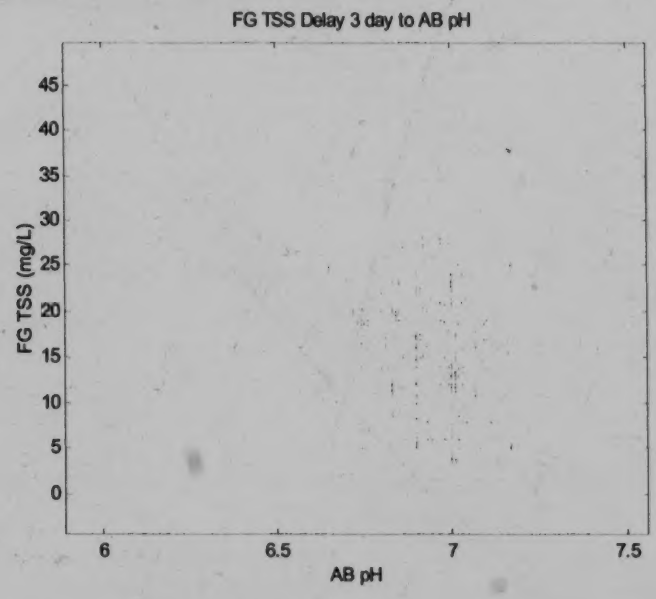


Figure 4.9 Scatter plot of AB pH and FG TSS, from day 108 to 419

4.3 Summary and Conclusions

The comprehensive data preparing process is a five-step process:

- Variables that are sampled more frequently than one day were converted into a sampling time of once per day;
- Missing data were added by interpolation.
- Possible trends of the data set were analyzed. It is confirmed the data set used for modeling is stationary.
- All means of the data set were removed. It's an essential procedure in linear prediction error modeling for the purpose of modeling dynamics of the process.
- The cross-correlation analysis indicates the degree of linear correlation between the inputs and output, FG TSS.

In general, the comprehensive data preprocessing prepared the data set for the modeling discussed in the following chapter. After synchronizing the sampling times, AB pH, AB Settleability and SC Torque were found most correlated to FG TSS by cross-correlation.

CHAPTER 5

MODELING EXPERIMENTS

Once a model's form is specified, unknown model parameters can be evaluated so that the accuracy of the model in representing the data can be assessed. This parameter estimation procedure is usually referred to as fitting the model, that is, as minimizing some overall differences between the model's predictions and the measured data. However, the problem of fitting a model to a set of input-output data becomes complicated when the model's relation is not simple or involves multiple inputs and outputs. Statistical analysis can be used to fit unknown model parameters and to evaluate the disturbance associated with the fitted model. In terms of the algorithms used to obtain the model parameters, the least square method (defines the estimate of the parameters as the values which minimize the sum of the squares of differences between the measurements and the predicted values) is commonly used in linear prediction error modeling.

The system identification problem is to estimate the parameters of a system based on the observed input-output data. In practice, the data set has to be separated into training, validation and testing set. The root mean square error (RMSE), a

measurement of the prediction error, used as the performance index of the models, is calculated as follows,

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(Y^{R}_{i} - Y^{M}_{i})^{2}}{N}}$$

where, Y^{R}_{i} is real value (measured),

Y^{M}_{i} is the predicted value by empirical model, and

N is the total number of data.

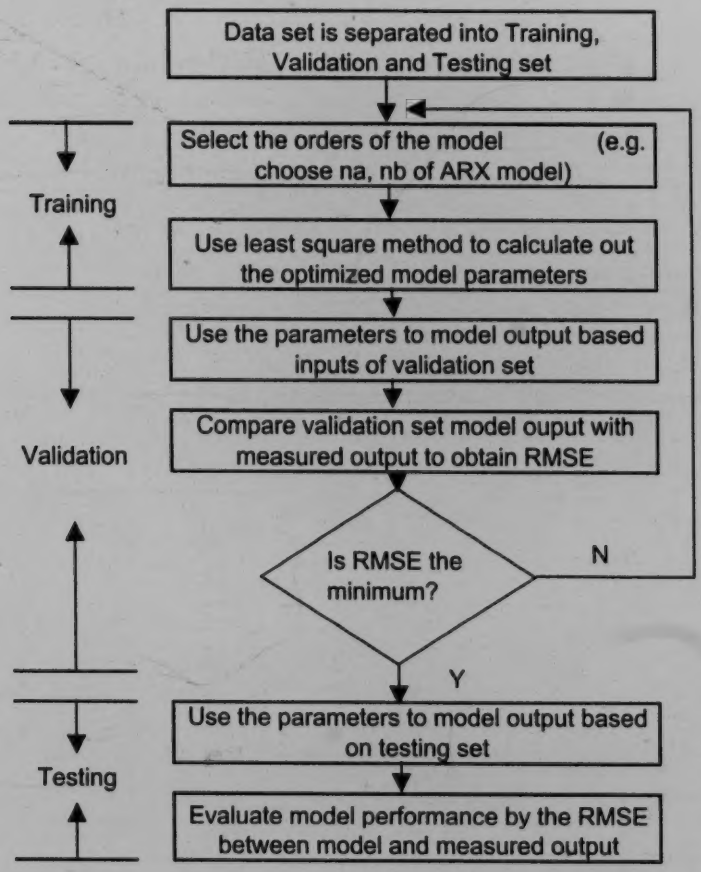

Data set is separated into Training, Validation and Testing set
Select the orders of the model (e.g. choose na, nb of ARX model)
Training
Use least square method to calculate out the optimized model parameters
Use the parameters to model output based inputs of validation set
Compare validation set model ouput with measured output to obtain RMSE
Validation
Is RMSE the minimum?
N
Y
Use the parameters to model output based on testing set
Testing
Evaluate model performance by the RMSE between model and measured output

Figure 5.1 The flowchart of linear prediction error model optimization

As shown in Figure 5.1, in the process of training, the parameters and order of the models are found by the following calculations. Using an ARX model as an example, na, nb are the order of the respective polynomials in the ARX model: $y\ (t) + a_1 y(t-1) + ... + a_{na} y(t-na) = b_1 u\ (t-1) + ... + b_{nb} u\ (t-nb) + e(t)$ (3.9)].

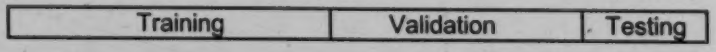
First, let na=1, nb=1 in the first round of searching. The least square method is used to calculate the optimized (i.e. minimizing RMSE between predicted and actual values) parameters for the linear prediction error model based on the set of input and output data in the training set with na=1, nb=1. Using the parameters gained from the training set, the validation set is used to output the model's prediction, and the performance index such as RMSE is stored as the minimum RMSE, say Min_RMSE.

Second, let na=1, nb=2, use the same inputs as first round in training to get the optimum parameters with least RMSE in training set, then the RMSE of validation set based on na=1, nb=2 is compared with the Min_RMSE. The smaller RMSE will replace the Min_RMSE.

Again, na=1, nb=3 is tried (put through the same process of training, validation and getting the smallest RMSE, stored in Min_RMSE). After na=1, nb=15 is tried, na=2, nb=1, 2,..,15 is tried again until na=15, nb=15 is tried. In this thesis, the maximum order of all models to be tried is 15. For example, the maximum order of ARX model is na=15, nb=15.

On the basis of the above analysis (cycles of training and validation), a model with the smallest RMSE is selected for further analysis.

The testing set is independent from the optimization process that is used to verify the performance of the model. That is, based on the model parameters obtained from the training and validation cycles, an independent set of inputs is put through the model, and the model's outputs and the measured outputs are compared. RMSE is used to indicate the difference, which is called the performance index of the model. We normally separate the data into training, validation and testing sets as follows.



As a rule of thumb, about or more than a half of the data set is used for training in order to learn the model's dynamics as much as possible. The validation set should have enough length to validate and optimize the model's performance. The testing set should be long enough to verify the model's prediction accuracy.

In Chapter 4, AB pH, AB Settleability and SC Torque are the highest correlated inputs to FG TSS. Four linear prediction error models, ARX, ARMAX, OE, BJ, discussed in Chapter 3, are applied to predict the FG TSS by using conventional data and by adding the respirometry data respectively. The searching of suitable inputs for prediction of FG TSS is conducted by comparing the prediction performances by possible inputs, even those, less correlated to FG TSS. From the comprehensive experiments, three typical examples of using less correlated inputs are shown below. First, using CC

NH_3, Temperature and AB pH as inputs, shown in Figure 5.2, the prediction of FG TSS deviates significantly from the measured FG TSS.

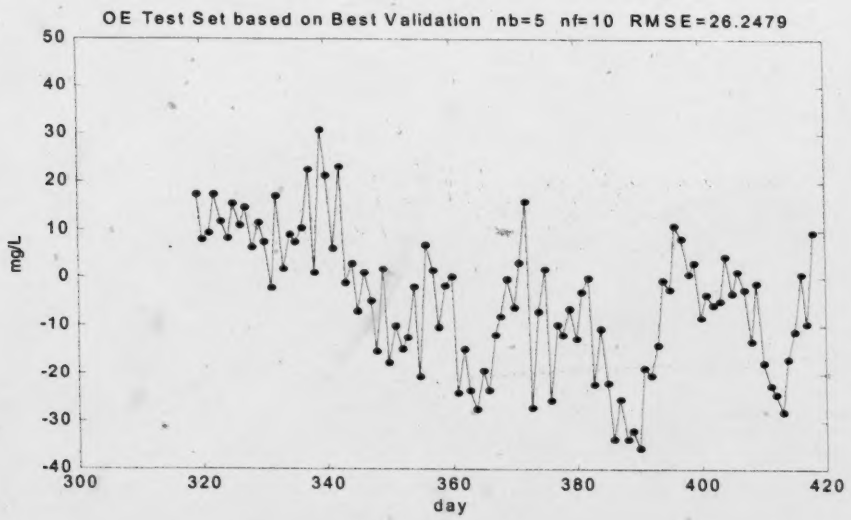


Figure 5.2 Using CC NH_3, Temperature, AB pH as inputs

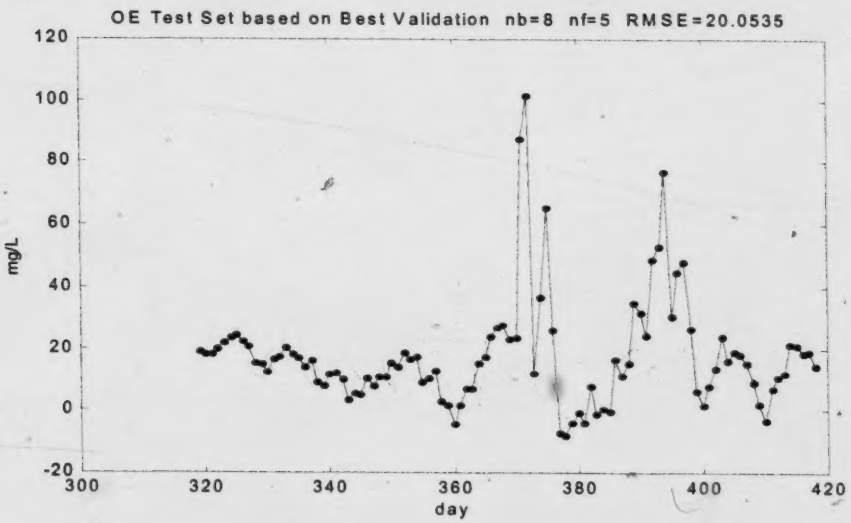


Figure 5.3 Using NH_3 Flow, P Flow, Temperature as inputs

Similarly, as shown in Figure 5.3, using NH_3 Flow, P Flow and Temperature as inputs, the prediction deviates quite significantly around the 370th and 390th day. This is likely because that NH_3 Flow and P Flow have several major changes around the 370th and 390th day in the whole data set sampling period. For the plots of the inputs, refer to APPENDIX 1.

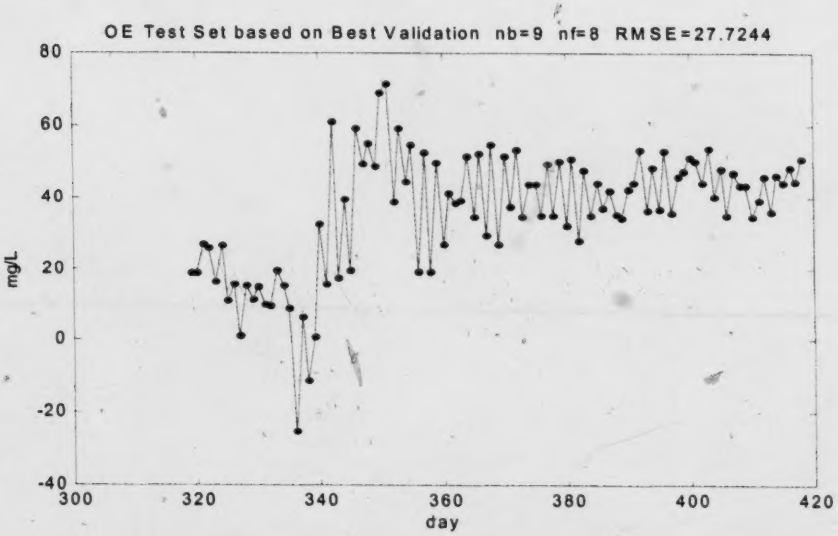


Figure 5.4 Using CC NH_3, WAS Flow, RAS Flow as inputs

Using CC NH_3, WAS Flow and RAS Flow as inputs, Figure 5.4 shows that the prediction shifts up after the 340th day. As a manipulated variable, the level of RAS Flow shifts up significantly after the 340th day, which causes the model output shift up correspondingly. For the plot of RAS Flow, refer to APPENDIX 1.

After narrowing down the inputs that can relatively better predict FG TSS, one-input, two-input and three-input prediction of FG TSS are conducted respectively. With AB pH, AB Settleability and OUR Shape change as three inputs, the OE model gives the best prediction of FG TSS. The linear system identification optimization (to find out the best coefficients and orders of the linear prediction error model) is conducted in the steps illustrated in Table 5.0.

Table 5.0 Experiments to be conducted by using different inputs

Step 1	**Using conventional inputs**
One-input	AB pH
Two-input	AB pH, Settleability
Three-input	AB pH, Settleability, SC Torque
Step 2	**Adding OUR shape for modeling**
One-input	OUR Shape Change
Two-input	AB pH, OUR Shape Change
Three-input	AB pH, Settleability, OUR Shape Change

As shown in Table 5.0, using the conventional process variables such as AB pH, Settleability and SC Torque as inputs, the first step is conducted by following three sub-steps: using AB pH as one input, using AB pH and settleability as two inputs, and using AB pH, Settleability and SC Torque as three inputs to predict FG TSS. The second step is conducted by adding the OUR shape change as another input to enhance the prediction. Finally, to evaluate the models by comparing their performance index RMSE and the prediction curves with the measured output. In the following experiments, the data set is separated into training, validation and testing sets

representing 65%, 17% and 18%, respectively, of the total data. To investigate the partitioning of the data set, another way of partitioning the three data sets is also explored. For example, we may use less data for training and more data for testing, which is presented in APPENDIX 2 and had no significant change in these results.

5.1 Using the conventional process data to predict FG TSS

AB pH, Setteleability and SC Torque were found to be the most correlated inputs to the output, FG TSS, They were used to predict FG TSS.

5.1.1 One-input (AB pH) case

Using AB pH as one input to predict FG TSS yields the following results.

Table 5.1 Performance index / model parameters (one input: AB pH)

Output: FG TSS mean = 15.98		RMSE Training	Validation	Testing	na	nb	nf	nc	nd
		Day 113~ 312	Day 311~362	Day 363 ~419					
One Input									
Input 1:AB pH	ARX	7.52	6.76	7.43	1	5			
correlation	ARMAX	7.49	6.68	7.53	2	3		7	
-0.482	OE	7.48	6.72	7.49		5	4		
FG TSS Delays 3 days to AB pH (nk=3)	BJ	8.07	6.43	8.03		7	13	8	6

As shown by Table 5.1, with different combinations of polynomial orders, na, nb, nf, nc and nd, the linear prediction error models give similar results with RMSE around

7.6, while the mean value of FG TSS is 15.98. The error is significant (7.6/15.98 ≈ 0.48). Using OE model as an example, its model structure is

$$y(t) = \frac{B(q)}{F(q)} u(t - nk)$$

where $B(q) = b_1 q^{-1} + \cdots b_{nb} q^{-nb}$

$F(q) = 1 + f_1 q^{-1} + \cdots f_{nf} q^{-nf}$

As indicated by Table 5.1, using nk=3, and when nb = 5, nf =4,

$$B(q) = b_1 q^{-1} + b_2 q^{-2} + b_3 q^{-3} + b_4 q^{-4} + b_5 q^{-5}$$

$$F(q) = 1 + f_1 q^{-1} + f_2 q^{-2} + f_3 q^{-3} + f_4 q^{-4}$$

Therefore,

$$F(q)y(t) = B(q)u(t)$$

$$y(t) + f_1 y(t-1) + f_2 y(t-2) + f_3 y(t-3) + f_4 y(t-4)$$
$$= b_1 u(t-4) + b_2 u(t-5) + b_3 u(t-6) + b_4 u(t-7) + b_5 u(t-8)$$

so the output,

$$y(t) = -f_1 y(t-1) - f_2 y(t-2) - f_3 y(t-3) - f_4 y(t-4)$$
$$+ b_1 u(t-4) + b_2 u(t-5) + b_3 u(t-6) + b_4 u(t-7) + b_5 u(t-8)$$

where, the coefficients of b_1 ,b_2 ,b_3 ,b_4, b_5 and f_1 ,f_2 ,f_3, f_4 are calculated by minimizing the linear prediction error. The model indicates that the current y(t) is affected by u(t-4), u(t-5), u(t-6), u(t-7)and u(t-8) as well as by y(t-1), y(t-2), y(t-3), and y(t-4).

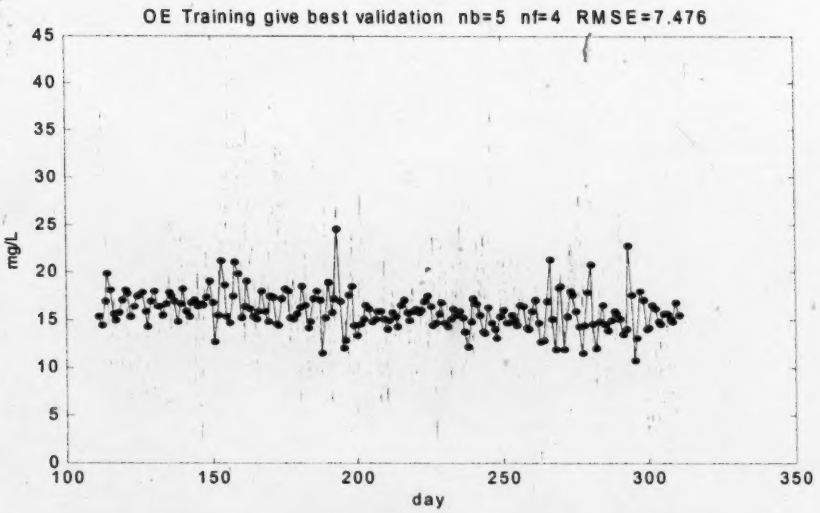


Figure 5.5 OE Training set, one input: AB pH

Figure 5.5 shows the training set of the OE modeling. The model output looks like a moving average of measured value. As shown by Figure 5.6, based on the parameters obtained from the training set, the linear model structure that gave the minimum RMSE for the validation data set was selected. Both the training and validation model curves are like an average trend of the measured values.

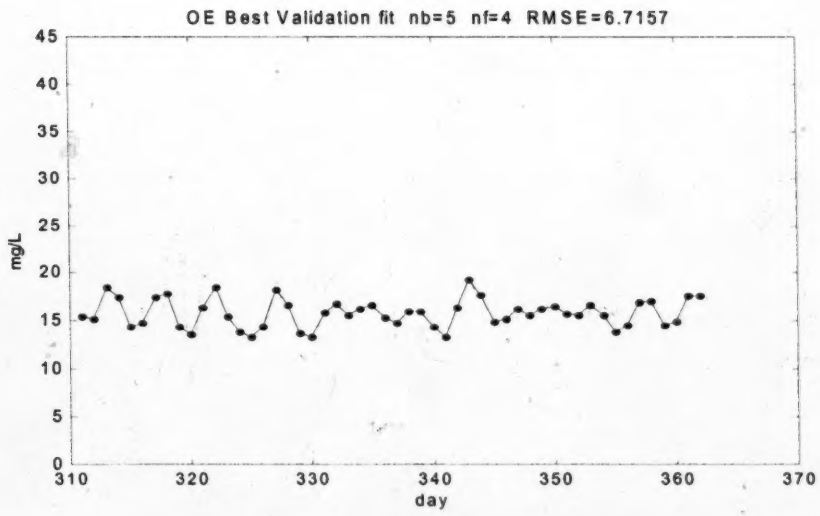


Figure 5.6 OE Validation set, one input: AB pH

After using the above optimization approach, an independent data set is used to test the prediction ability of this OE model. As shown in Figure 5.7, it cannot predict the change of FG TSS around the mean. Especially, from day of 410 to 419, the model output cannot predict a relatively major change of measured data. Other models should be found to improve the performance.

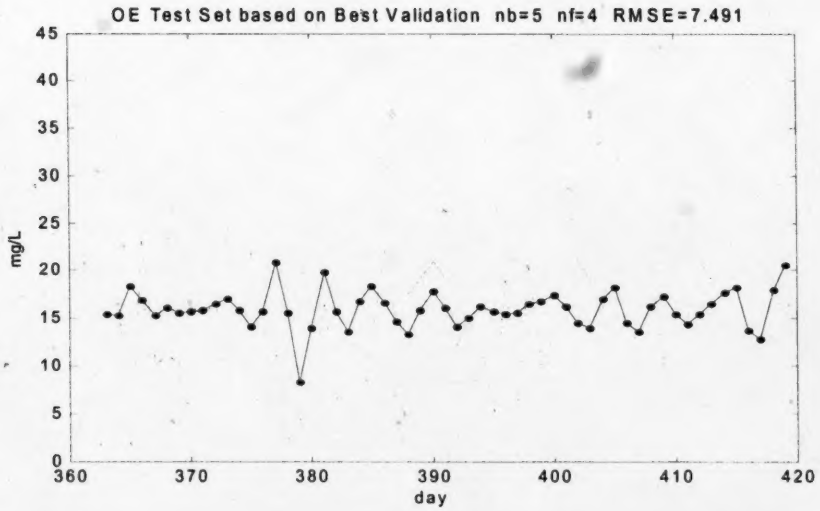


Figure 5.7 OE testing set, one input: AB pH

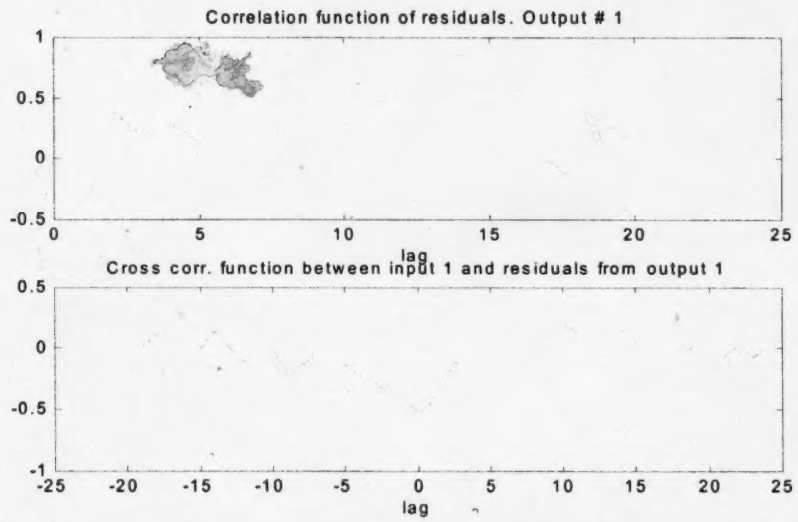


Figure 5.8 OE testing set residual analysis

Applying the theory presented in Section 3.1.2, Figure 5.8 shows the auto-correlation of the residuals (= the measured value – the predicted values of the testing set), and

the input (AB pH)'s cross-correlation with the residuals for the testing set. The AB pH's correlation with residual falls out of the 99% confidence limits (the dotted line). It means we do not have 99% of confidence that the residuals are uncorrelated with the input. Therefore, we cannot use the model without the noise term, e(t). Further work is needed to obtain a better model.

5.1.2 Two-input (AB pH, Settleability) case

Using AB pH and Settleability as inputs to predict FG TSS yielded the following results.

Table 5.2 Performance index / model parameters (two inputs: AB pH, Settleability)

Output: FG TSS mean = **15.98**		RMSE			na	nb	nf	nc	nd
		Training	Validation	Testing					
		Day 113~ 312	Day 313~362	Day 363 ~419					
Input 1:AB pH	ARX	7.14	7.18	7.03	8	9			
	ARMAX	7.13	7.22	6.96	5	5		9	
Input 2: Settleability	OE	7.19	7.24	6.72		4	2		
FG TSS Delay 3 days to Setteability	BJ	7.22	7	7.4		6	9	9	9
FG TSS Dealy 5 days to Settleability	OE	7.03	6.87	6.82		4	2		

Table 5.2 shows that the linear prediction error models give slightly better results compared with those of the one input case (Table 5.1). The OE, overall, gives a relatively better performance than the others (ARX, ARMAX, BJ). The plot of FG TSS testing set as well as that of the OE model's prediction is shown in Figure 5.9. A

comparison of Figure 5.9 to Figure 5.7 shows that the prediction performance is improving. In particular, the two-input model output after day 410 gives better prediction. Table 5.2 shows that FG TSS delay to settleability 3 days and 5 days give similar result by OE model. Since the delay of 3 days is consistent with the designed retention time of the plant, the three-day delay is used in the following tests.

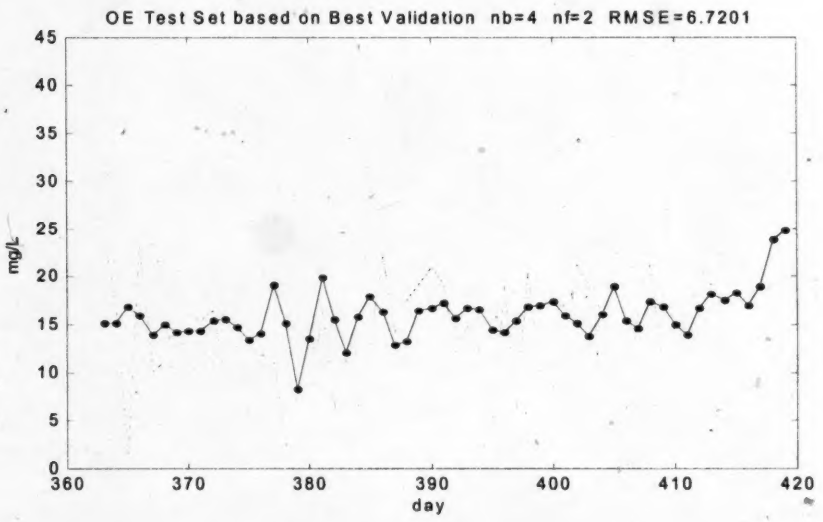


Figure 5.9 OE Testing set, inputs: AB pH, Settleability

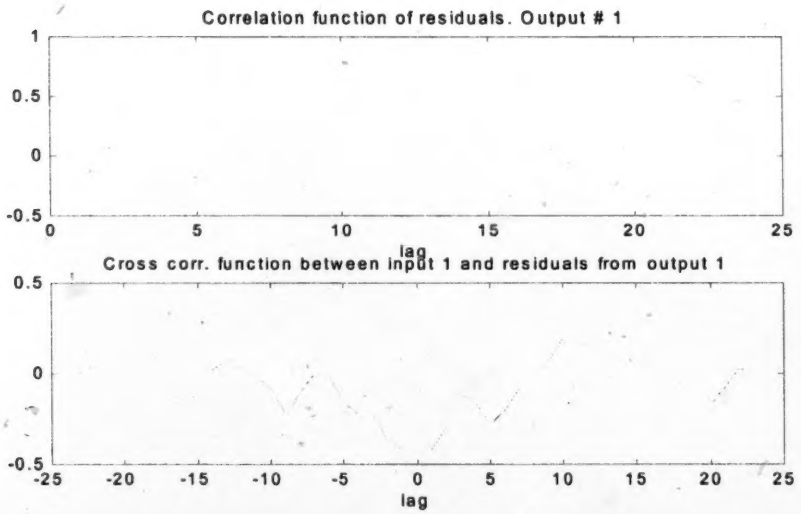


Figure 5.10 OE testing set residual autocorrelation, input 1's cross-correlation with the residuals

Corresponding to Figure 5.9, Figure 5.10 shows the auto-correlation of the residuals, and input 1 (AB pH) 's cross-correlation with the residuals. Figure 5.11 shows the cross-correlation between input 2 (Settleability) and the residuals. The residuals just fall into the 99% confidence limits (the dotted line).

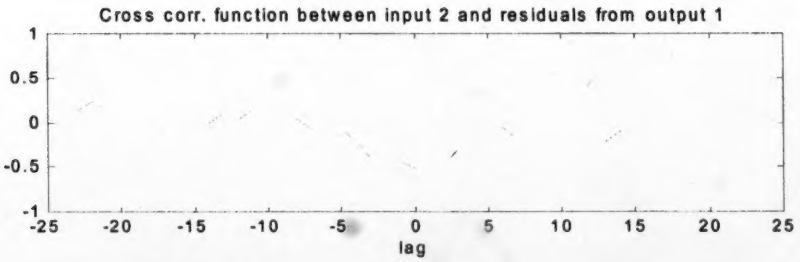


Figure 5.11 Input 2's cross-correlation with residuals

5.1.3 Three-input (AB pH, Settleability, SC Torque) case

In order to further improve the model's performance, three-input modeling is conducted. Based on the cross-correlation analysis in Section 4.2, SC Torque is the third most correlated variable to FG TSS. Therefore, AB pH, Settleability and SC Torque are used for modeling. The testing set of FG TSS versus the OE model's prediction is shown in Figure 5.12.

The adding of the third variable does not improve the performance. A comparison of Figure 5.12 with Figure 5.9 shows that the prediction is not improved in terms of predicting the change of FG TSS. Also, the RMSE of this three-input model is 7.36, while that of the two-input case is 6.72. Therefore, adding the 3rd input, SC Torque, does not improve the overall model's performance.

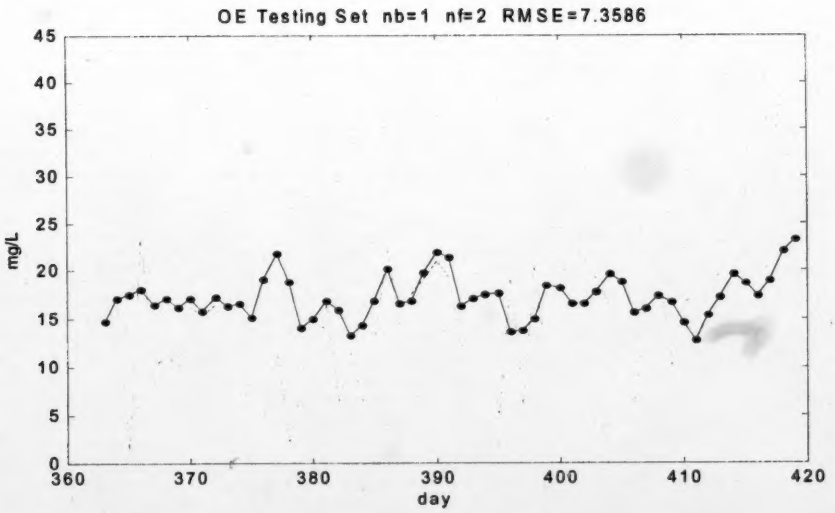


Figure 5.12 OE testing set, inputs: AB pH, Settleability, SC Torque

5.2 Using the Respirometer data to improve the prediction of FG TSS

The data from the respirometer (location is shown in Figure A1.3) can provide "snap-shots" of the current AST process behavior. Particularly, the OUR (oxygen uptake rate) curve is quite sensitive to the biomass's activity. On the basis of the discussion in Section 1.4, the analysis of OUR curve is conducted as follows.

5.2.1 General analysis of OUR curve

Figure 5.13 shows the results of a typical batch measurement. When one batch enters the respirometer, with sufficient food and nutrients, from hour 1.1 to hour 1.5 (i.e., a 0.4-hour duration), OUR reaches the maximum value very quickly. Once the OUR reaches the maximum, it drops quickly for a period of 1 hour (from hour 1.5 to 2.5). The biomass is quite active during the first 1.4 hours and quickly takes in food as well as nutrients. Then the OUR curve's slope becomes less steep, gradually declining from hour 2.5 to hour 6, while most of the food and nutrients (NH3 and P) are consumed. After hour 6, when NH_3 is gone, the biomass enters the endogenous period. In this endogenous period, the OUR curve's slope becomes even less steep than the period from hour 2.5 to hour 6. With minimum activity of biomass, the P level is relatively constant in this endogenous period.

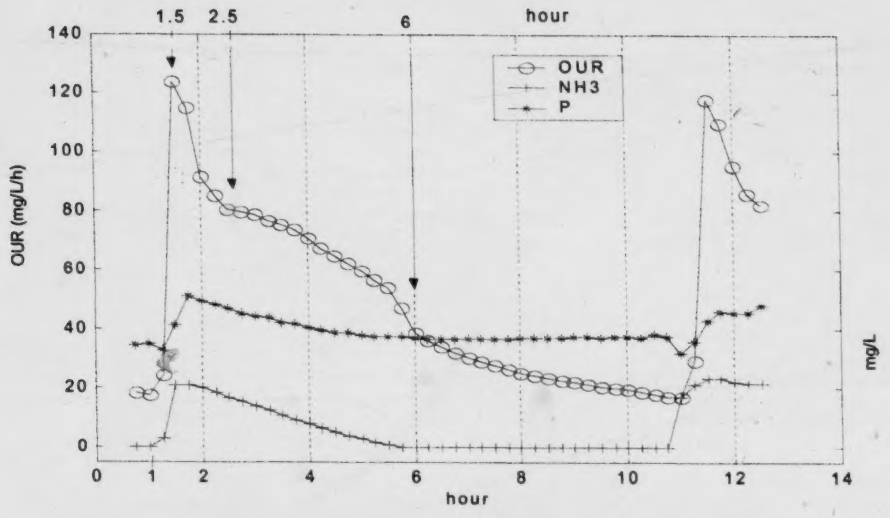


Figure 5.13 One cycle of OUR and Nutrients Change

As shown in Figure 1.1, the ARC respirometer takes a batch (as shown in Figure 1.1) from the contact chamber (1 to 4 times a day) just before the effluent enters the activated sludge basin and runs like a pilot plant of the activated sludge basin and thus provides "snap-shots" of the current AST process behavior.

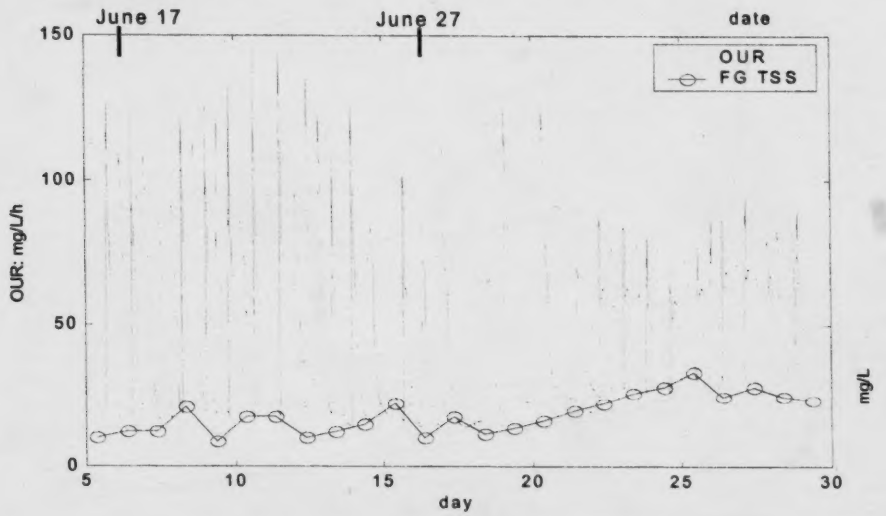


Figure 5.14 OUR cycles and FG TSS

FG TSS is plotted in Figure 5.14 against OUR from the 6th day to the 29th day. The respirometer starts running at ANC on June 11. The period from June 17 (day 7) to June 27 (day 17) is considered a relatively stable period for the WWT plant. The OUR values during this period may be used to develop a benchmark to compare with other OUR curves. Each OUR curve indicates each batch collected and analyzed by the respirometer. The OUR shape is quite different during the period when FG TSS is increasing (after day 22). This feature of the OUR curve is analyzed in the next section.

5.2.2 Feature extraction of OUR curves

To develop algorithms for outputting respirometer variables, some of the respirometer data may require further processing into new process variables. For example, the shapes of the respiration curves are likely to be sensitive indicators of biomass health and as such, may contain considerable information useful for process diagnosis and control.

In Figure 5.14, during the period shown from June 17 to June 27, the FG TSS is relatively constant when compared with that in the later period shown. All the OUR curves have similar shapes from June 17 to June 27. Hence, the OUR curves in this period were considered "normal", and an ensemble average of 11 out of these curves is shown in Figure 5.15.

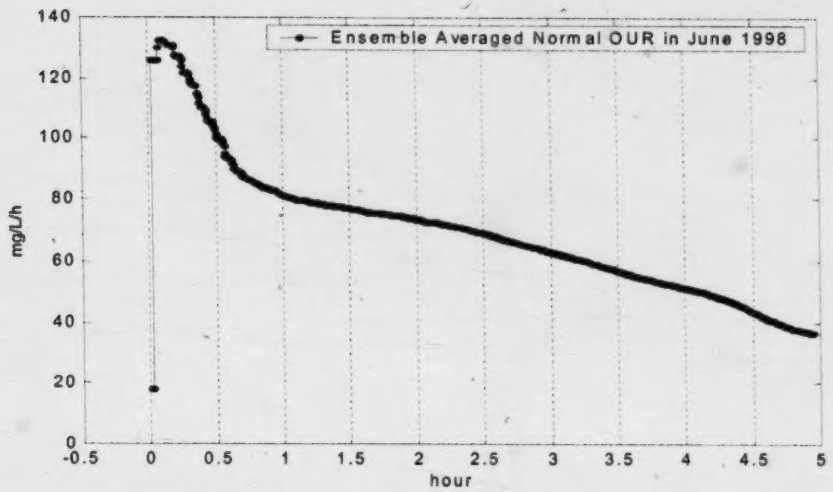


Figure 5.15 Ensemble Averaged OUR in June 1998

For subsequent times, "non-normal" OUR shapes were observed. One typical of them is shown in Figures 5.16.

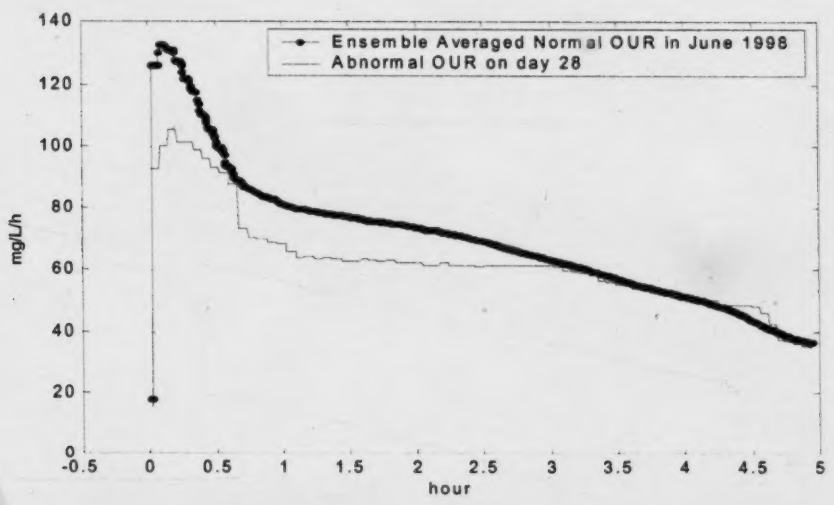


Figure 5.16 Ensemble Averaged OUR vs. OUR on 28th day

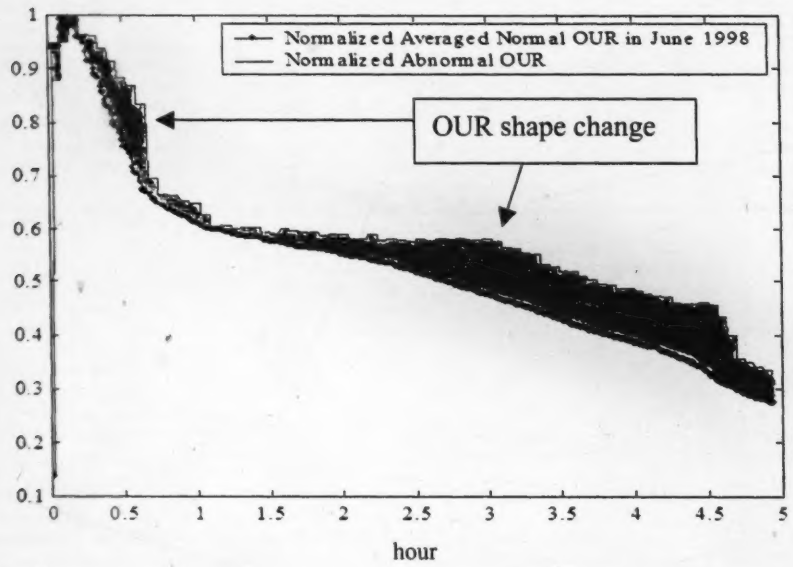


Figure 5.17 Averaged OUR vs. OUR on 28th day (all are normalized)

Figure 5.16 reveals that a "bumpy" OUR shape occurred on the 28th day. This shape is considered to be "abnormal" when the OUR increases in the middle of the batch cycle instead of decreasing smoothly. This OUR curve is compared with the averaged OUR. To capture the change of the OUR curve's shape rather than its magnitude, the bumpy curve and the averaged OUR curves are normalized (divided by the OUR curve's maximum value) to the range of 0 to 1 and shown in Figure 5.17. We wish to quantitatively characterize this abnormal behavior and then determine if it can be used in the prediction of FG TSS. Shown in Figure 5.17, the absolute value of the sum of the OUR shape's differences (value difference between each OUR value and the value of the ensemble averaged OUR) was chosen as condensed information about the OUR curve to represent the features of the batch. This absolute value could be correlated

with other conventional AST data obtained by the mill staff once per day. The difference is then calculated as the magnitude of OUR minus the normal ensemble averaged OUR, and the sum of the absolute difference, also called the "shape change", is thus developed. This is shown in Figure 5.18. The OUR shape change from 108th day to 419th day are used for the following modeling. Using the OUR shape change from 1st day to 419th to predict FG TSS is presented in APPENDIX 3.

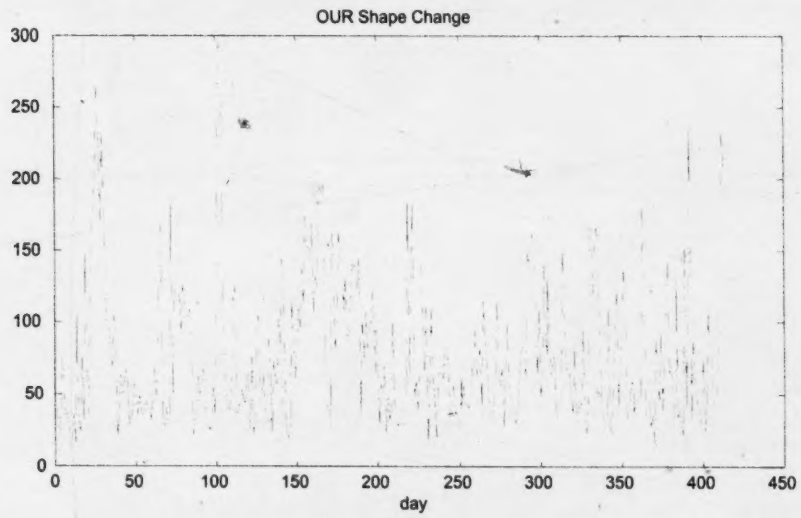


Figure 5.18 OUR shape change

As shown in Figure 5.18, the value of OUR Shape Change surges around the 25th day, which is the same period as shown in Figure 5.14. Indeed, the OUR Shape Change can capture the shape change of OUR curve.

5.2.3 One-input (OUR Shape change) case

Using the OUR shape change alone as an input to predict FG TSS, shown in Table 5.3, the performance index is similar to that of one conventional input case (Table 5.1) in terms of the absolute RMSE value.

Table 5.3 Performance index /model parameters (one input: OUR shape change)

Output: FG TSS mean = **15.98**		RMSE			na	nb	nf	nc	nd
		Training	Validation	Testing					
		Day 113~ 312	Day 313~362	Day 363 ~419					
Input 1:	ARX	7.5	6.96	7.58	4	8			
OUR shape change	ARMAX	7.56	6.9	7.8	1	7		2	
FG TSS Delays 5 days to input 1	OE	7.71	6.92	7.87		4	7		
	BJ	7.57	6.59	7.75		6	10	7	6
OUR shape change FG TSS Delays 3 days to input 1	OE	7.68	6.86	7.88		1	7		

The plot of OE model testing set is shown in Figure 5.19. OUR shape change alone cannot provide enough information for the prediction of changes in FG TSS. Further experiments are necessary to better predict FG TSS, especially after the 410^{th} day. The RMSE when FG TSS delays to OUR shape change 3 days (with slightly higher cross-correlation coefficient) or 5 days are similar. Since the delay of 5 days is consistent with designed retention time of the plant, the 5-day delay is used in the following tests.

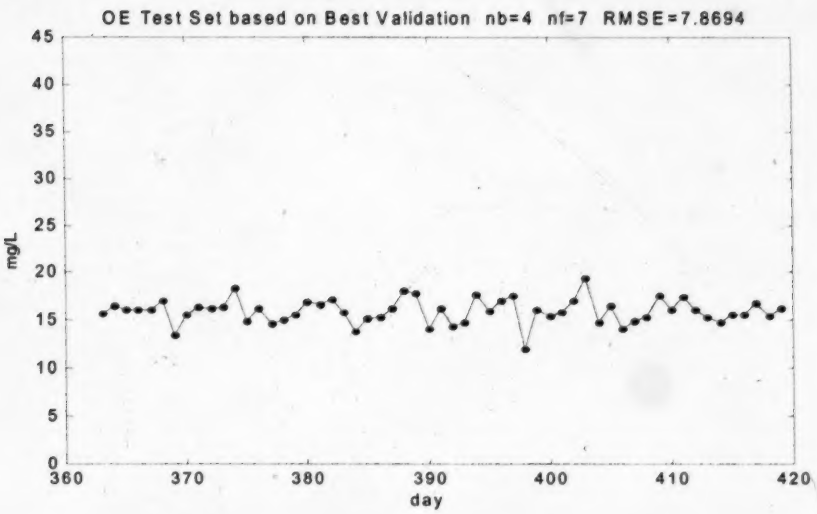


Figure 5.19 OE testing set, one input: OUR shape change

5.2.4 Two-input (AB pH & OUR Shape change) case

Using AB pH and OUR shape change as inputs to predict FG TSS, shown in Table 5.4, the performance index is similar in terms of absolute RMSE value to that of the previous case.

Table 5.4 Performance index /model parameters (two inputs: AB pH, OUR shape change)

Output: FG TSS mean = 15.98		RMSE			na	nb	nf	nc	nd
		Training	Validation	Testing					
		Day 113~ 312	Day 313~362	Day 363 ~419					
Input 1:AB pH	ARX	7.5	6.92	7.73	2	10			
	ARMAX	7.28	7	8.2	5	10		7	
Input 2:	OE	7.2	7.06	7.42		3	5		
OUR shape change	BJ	7.71	6.86	8.45		6	8	6	7
FG TSS Delays 5 days to input 2									

The plot of OE model's testing set is shown in Figure 5.20. Comparing with the results shown in Figure 5.19, the results in Figure 5.20 show better prediction of FG TSS. It starts to predict the major change after day 410. However, the prediction is more or less like a straight line (not predicting of dynamics), further experiments are necessary to better predict FG TSS.

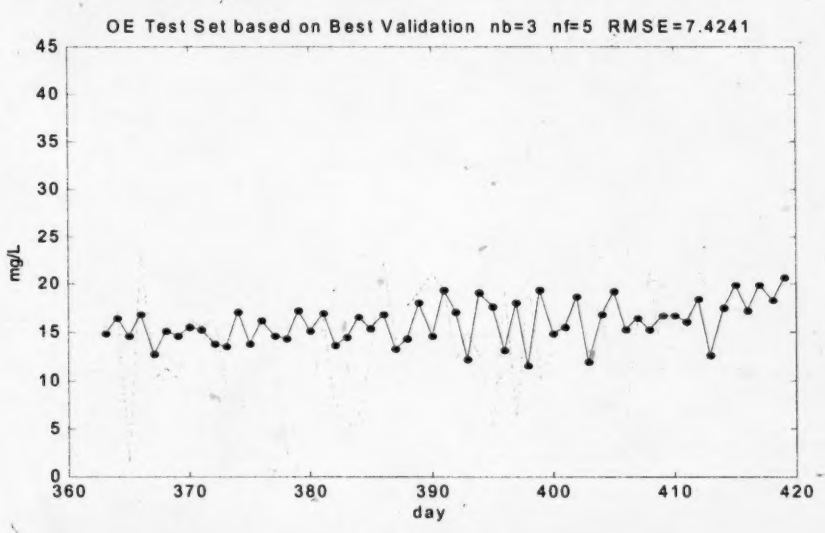


Figure 5.20 OE testing set, two inputs: AB pH, OUR shape change

5.2.5 Three-input (AB pH, Settleability & OUR Shape change) case

As shown in the performance index (RMSE) Table 5.5, this case is similar to the two-input case (using two conventional variables; refer to Table 5.2) in terms of absolute RMSE value.

Table 5.5 Performance index /model parameters (three inputs: AB pH, Settleability, OUR shape change)

Output: FG TSS mean = **15.98**		RMSE Training	Validation	Testing	na	nb	nf	nc	nd
		Day 113~ 312	Day 313~362	Day 363 ~419					
Input 1:AB pH	ARX	7.25	7	7.21	2	3			
	ARMAX	6.91	7	7.25	5	5		5	
Input 2: Settleability	OE	6.7	6.9	6.89		5	9		
	BJ	7.6	6.47	7.61		5	9	9	6
Input 3: OUR Shape Change									

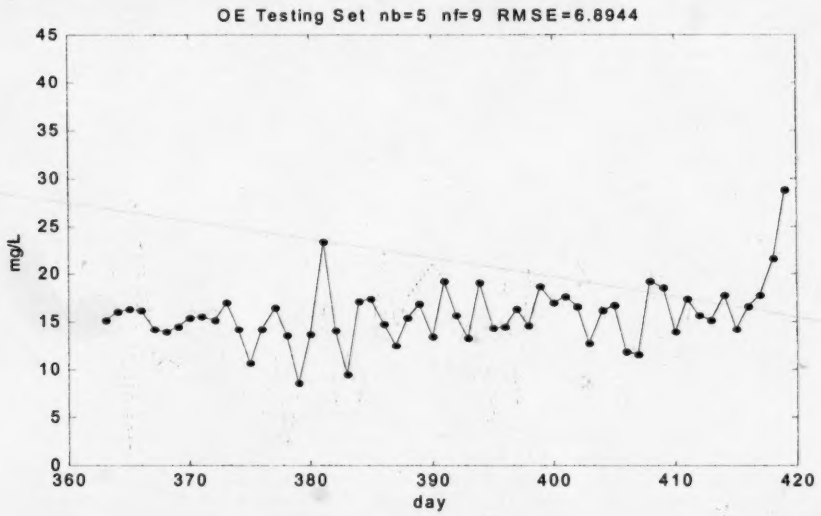


Figure 5.21 OE testing set, three inputs: AB pH, Settleability, OUR shape change

To evaluate the performance of the models, the absolute value of RMSE should not be the only criteria. As shown in Figure 5.21, although the absolute RMSE value is similar to that obtained by using two conventional inputs (Figure 5.9), the prediction of FG TSS change, is improved. Especially, on day 419, the two-input model's prediction is 25, the three-input model's prediction is 29, while the measured value is 36. Since the prediction of major change of FG TSS and to take action to correct the

change in time is the objective of WWT process operation, this three-input model (adding OUR Shape change as the 3rd input) is more useful.

By comparing Figure 5.21 with Figures 5.7, 5.9, 5.12, 5.19 and 5.20, we can conclude that using the three inputs of AB pH, Settleability and OUR shape change gives the best prediction of the change of FG TSS. Using the OE model can help the operators to take proper process control action three days in advance to avoid the mistreated effluent being discharged into the receiving river.

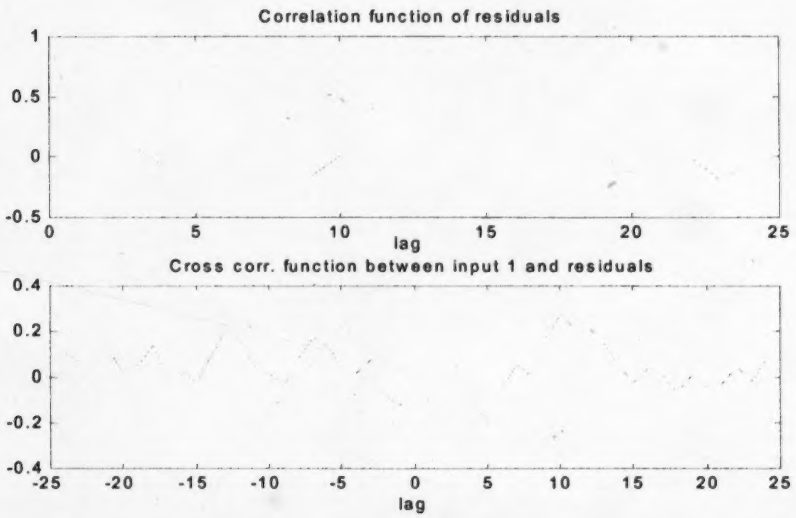


Figure 5.22 Three-input OE testing set residual autocorrelation, input 1's cross-correlation with the residuals

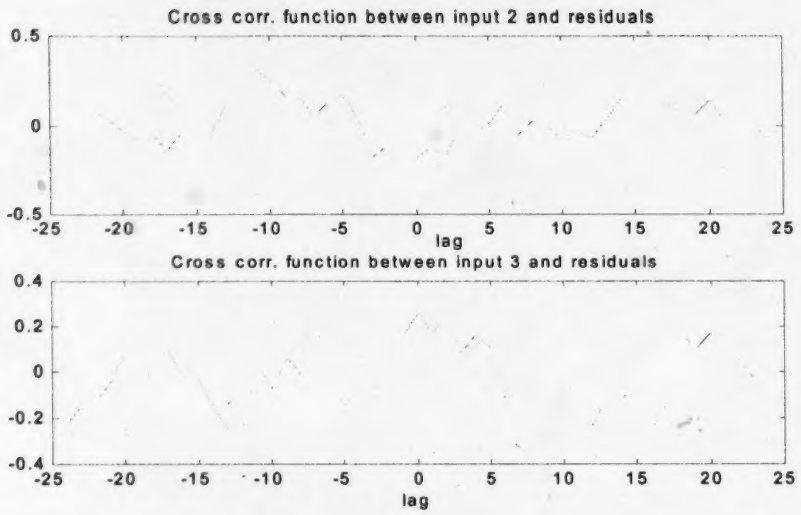


Figure 5.23 Three-input OE testing set, input 2's and input 3's cross-correlation with the residuals

As shown in Figures 5.22 and 5.23, the residuals are not correlated much with the inputs and between themselves. They all fully fall into the 99% confidence limits (the dotted line). Thus, we can treat the residuals as white noise and use the model to predict FG TSS without including the noise term, e(t).

The training set for the three-input case is presented in Figures 5.24. As shown, the model mostly fits the training set quite well. Comparing Figure 5.24 with Figure 5.5, the training set of three-input fits the measured value much closer. It means the three-input model can learn more of the process than one-input model does. Certainly, by

better learning the process, the model can give better performance in predicting the process changes.

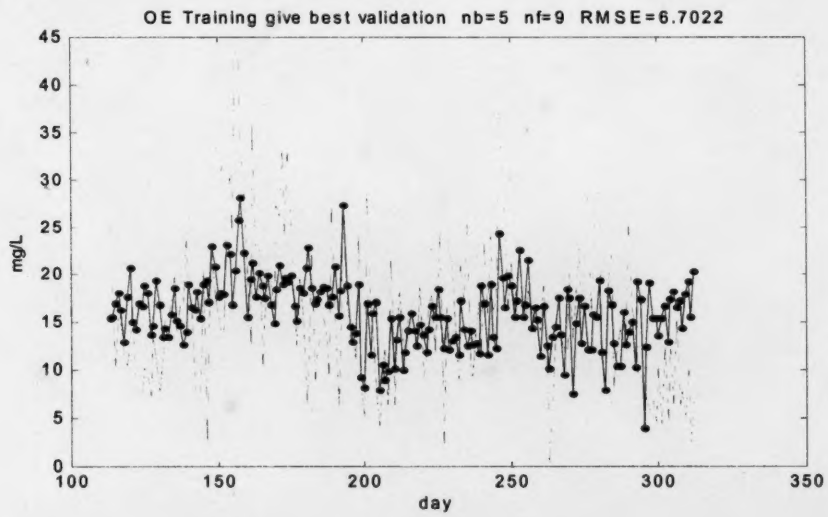


Figure 5.24 The training set of OE model, three inputs: AB pH, Settleability, OUR shape change

5.3 Conclusion of model performance and its application

5.3.1 Comparison of modeling performance

As shown in Table 5.6, comparing all the testing sets of the above models, among Figures 5.7, 5.9, 5.12, 5.19, 5.20 and 5.21, the linear prediction error model with AB pH and Settleability as two-input (Figure 5.9) gives the lowest RMSE. However, if we compare Figure 5.21 and 5.9, with slightly higher RMSE, the prediction obtained by

three-input – AB pH, Settleability and OUR Shape change is better for predicting the changes, especially major changes, of FG TSS. Since the prediction of major changes of FG TSS is the valuable information for plant operation, the OE model of the three-input case is thus selected for predicting the change of FG TSS.

Table 5.6 Summary of the performance (Predicted vs. Measured FG TSS) of models

Inputs	Predicted vs. Measured FG TSS	Reference figure
Linear Prediction Error Modeling	OE model RMSE	
AB pH	7.49	Figure 5.7
AB pH, Settleability	6.72	Figure 5.9
AB pH, Settleability, SC Torque	7.36	Figure 5.12
OUR Shape Change	7.87	Figure 5.19
AB pH, OUR Shape Change	7.42	Figure 5.20
AB pH, Settleability, OUR Shape Change	6.89	Figure 5.21

Static multi-layer perceptron neural networks have also been used in attempts to predict the process output, FG TSS. The results are not better than the above models and the results are summarized in APPENDIX 4.

5.3.2 Application of the model

Based on the above analysis, an OE model using the inputs of AB pH, Settleability and OUR shape change, gives an applicable prediction of FG TSS in advance of the treated effluent's discharge into the receiving river. The detailed description of the application of this model is illustrated below.

According to equation (3.12), the structure of OE model with multiple inputs is

$$y(t) = \frac{\mathbf{B(q)}}{\mathbf{F(q)}}\mathbf{u(t-nk)}$$

where $\mathbf{B(q)} = B_1q^{-1} + \cdots B_{nb}q^{-nb}$, $\mathbf{F(q)} = 1 + F_1q^{-1} + \cdots F_{nf}q^{-nf}$, **nk** is the output delay to inputs.

Refer to Table 5.5, the optimized parameters of the three-input OE model are nb=5, nf=9; thus,

$$\mathbf{B(q)} = B_1q^{-1} + B_2q^{-2} + B_3q^{-3} + B_4q^{-4} + B_5q^{-5}$$

$$\mathbf{F(q)} = 1 + F_1q^{-1} + F_2q^{-2} + F_3q^{-3} + F_4q^{-4} + F_5q^{-5} + F_6q^{-6} + F_5q^{-7} + F_5q^{-8} + F_5q^{-9}$$

Therefore, y(t), the prediction of measured value, is expressed as follows,

$$y(t) = \frac{\mathbf{B(q)}}{\mathbf{F(q)}}\begin{pmatrix} u_1(t-3) \\ u_2(t-3) \\ u_3(t-5) \end{pmatrix} \qquad (5.1)$$

The coefficients **B(q)** and **F(q)** can be expressed in the matrix form, where the coefficients are calculated by the optimization of the linear prediction error performance. That is,

$$\mathbf{F} = \begin{bmatrix} 1 & f_1^1 & f_2^1 & f_3^1 & f_4^1 & f_5^1 & f_6^1 & f_7^1 & f_8^1 & f_9^1 \\ 1 & f_1^2 & f_2^2 & f_3^2 & f_4^2 & f_5^2 & f_6^2 & f_7^2 & f_8^2 & f_9^2 \\ 1 & f_1^3 & f_2^3 & f_3^3 & f_4^3 & f_5^3 & f_6^3 & f_7^3 & f_8^3 & f_9^3 \end{bmatrix} =$$

$$\begin{pmatrix} 1.0000 & 1.0024 & 1.0427 & 0.4098 & 0.0660 & -0.5937 & -0.5255 & -0.3502 & -0.2815 & 0.0103 \\ 1.0000 & -0.0558 & 0.1408 & -0.5433 & -0.1440 & -0.2124 & -0.4835 & 0.1944 & 0.1039 & 0.4481 \\ 1.0000 & -1.2142 & 0.5642 & -0.3163 & 1.5635 & -1.3308 & 0.4063 & -0.5365 & 1.1490 & -0.6420 \end{pmatrix}$$

$$\mathbf{B} = \begin{bmatrix} b_1^1 & b_2^1 & b_3^1 & b_4^1 & b_5^1 \\ b_1^2 & b_2^2 & b_3^2 & b_4^2 & b_5^2 \\ b_1^3 & b_2^3 & b_3^3 & b_4^3 & b_5^3 \end{bmatrix} =$$

$$\begin{pmatrix} 1.8908 & 2.9103 & -6.8316 & -17.1563 & -8.8958 \\ -0.0059 & 0.0049 & -0.0117 & -0.0057 & 0.0121 \\ 0.0002 & -0.0040 & 0.0158 & -0.0148 & 0.0079 \end{pmatrix}$$

By inserting the values of the inputs into equation (5.1), we can calculate out the y(t) to predict FG TSS. For example, as shown in Figure 5.25, insert inputs into equation (5.1), using the AB pH (u_1), Settleability (u_2) value on day 3, and the OUR Shape change (u_3) value on day 1, that is,

$$y(6) = \frac{\mathbf{B(q)}}{\mathbf{F(q)}} \begin{pmatrix} u_1(6-3) \\ u_2(6-3) \\ u_3(6-5) \end{pmatrix} = \frac{\mathbf{B(q)}}{\mathbf{F(q)}} \begin{pmatrix} u_1(3) \\ u_2(3) \\ u_3(1) \end{pmatrix},$$

the operator can use y(6) to predict the FG TSS on day 6 (shown in Figure 5.26). Accordingly, we can use the data on day 2 and day 4 to predict the FG TSS on day 7. The data presented here are the data from day 1 to day 11 in the whole data set (day 1

to day 419). There is no major change during this period. A typical prediction of a major change happened after day 410 is shown in Figure 5.21.

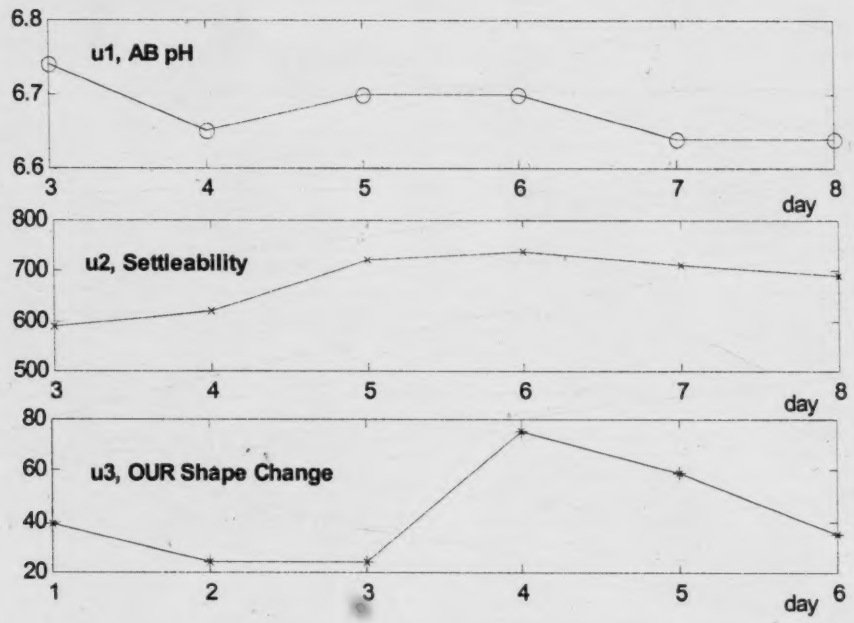


Figure 5.25 Inputs needed to predict the FG TSS

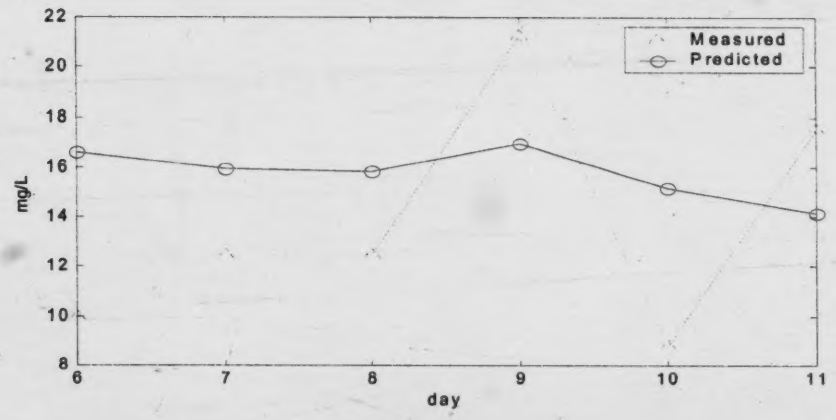


Figure 5.26 Measured and predicted value of FG TSS

In general, the model can predict the FG TSS three days in advance. While monitoring the variables measured before the effluent reaches the secondary clarifier, we can predict the FG TSS that will be measured 3 days later. This prediction can allow the operators to take proper action to reduce the FG TSS, and thus to prevent harmful effluent from being discharged into the receiving river.

CHAPTER 6

CONCLUSIONS AND RECOMMENDATIONS

The results obtained from the experiments suggest that the following objectives of this research have been achieved.

- Linear prediction error models can be applied to predict a key effluent output in a real pulp mill WWT system.
- Of the models using conventional process variables, the output error (OE) model with two selected inputs provided the best prediction (minimum root mean square error) of the final grab total suspended solids (FG TSS). FG TSS is one of the key process outputs and is measured just before the treated effluent being discharged into the river.
- A new variable has been extracted from one of the respirometer data, oxygen uptake rate (OUR). Adding the new variable as the third input, the prediction performance of this OE model was improved.

The model as proposed holds some limitations. Most critical of these is the incompleteness of field data and the disturbances to the real WWT process.

Comprehensive techniques have been applied to preprocess the field data used for modeling. Based on the research, the above linear empirical modeling approaches can predict the change of FG TSS three-day in advance. This allows corrective actions to be taken to minimize the amount contaminants discharged into the river. Because the settings of WWT plants vary significantly by location, empirical modeling is likely one of the most practical and economical methods that can be used.

In future research, the following directions could be explored to achieve more accurate and economical pulp mill WWT process modeling:

- The available industrial data were used for the analysis in this thesis. If it were possible to have a more frequent sampling rate, other key process variables such as NH3, P and BOD5 could be used in the modeling.
- The static MLP neural networks did not improve the prediction of the FG TSS over the simple linear models. Since the biochemical process is nonlinear in nature, dynamic MLP neural networks could be explored and may produce better results.
- Since the OUR is quite sensitive to the biomass's activity, by collecting OUR data from both input and output of the WWT plant, it could reveal more information of the dynamic biochemical process.
- The OUR is a representative variable that can indicate the overall biomass activity. However, we need to install an expensive respirometer to obtain it. With carefully selected and well collected conventional process variables, we may be able to construct a variable that has the same or even better function than OUR does.

BIBLIOGRAPHY

Andrews, J.F. (1974). Dynamic models and control strategies for wastewater treatment processes. Water Research, Vol.8, pp. 261-289.

Andrews, J.F. (1989). Dynamics, Stability and Control of the Anaerobic Digestion Process. Dyanmics Modeling and Expert Systems in Wastewater Engineering, Lewis Publishers, Michigan, pp92

Bastin, G. and Dochain, D. (1990). On-line Estimation and Adaptive Control of Bioreactors. Elsevier, Amsterdam.

Bhat, N. and McAvoy,T. (1990). Use of neural nets for dynamic modeling and control of chemical processes systems. Computers and Chemical Engineering, Vol.14, pp. 573-583.

Byrkit, D. (1987). Statistics today, a comprehensive introduction. The Benjamin/Cummings Publishing Company, Inc., Menlo Park, CA

Hack, M. and Kohne, M. (1996). Estimation of wastewater process parameters using neural networks. Water Science and Technology, Vol. 33, No. 1, pp. 101-115.

Health Canada, Minister of Public Works and Government Services Canada (1997) PULP AND PAPER MILL EFFLUENTS http://www.hc-sc.gc.ca/ehp/ehd/catalogue/general/iyh/pulp.htm April 3, 2003

Hennig, K. et al (1999). Multi-sensor system for fast analysis in environmental monitoring with an application in wastewater treatment. Institute of Biophysics, University of Hanover. http://las.physik.uni-oldenburg.de/projekte/earsel/4th_workshop_paper/hennig.pdf Dec.01 2001

Henze, M. et al (1987). Activated sludge model no. 1. Scientific and Technical Report No. 1. IAWQ, London, Great Britain.

Holmberg, U., Olsson, G. and Andersson, B. (1989). Simultaneous DO control and respiration estimation. Water Science and Technology, Vol. 21, pp.1185-1195.

Hoskins, J.C. and Himmelblau, D.M. (1988). Artificial neural network models of knowledge representation in chemical engineering. Computers and Chemical Engineering, Vol.12, 881-890.

Johnson, R. (1939). Elementary Statistics. Duxbury Press, Belmont, CA

Lindberg, C.F. (1997). Control and Estimation Strategies Applied to the Activated Sludge Process. PhD Thesis, Uppsala University.

Lindberg, C.F. (1998). Multivariable modeling and control of an activated sludge process. Water Science and Technology , Vol. 37, No. 12, pp. 149-156.

Ljung, L. (1999). System identification –Theory for the user. 2nd ed, Prentice Hall, Upper Saddle River, N.J.

Lukasse, L.J.S., Keesman, K.J. and Straten G. V. (1996). Grey-box identication of dissolved oxygen dynamics in activated sludge processes.13th World Congress of IFAC, San Francisco

McAuley, D. The BackPropagation Network: Learning by Example http://www2.psy.uq.edu.au/~brainwav/Manual/BackProp.html#Introduction. September 01, 2003.

Minsky, M. and Papert, S. (1969). Perceptrons: An Introduction to Computational Geometry. MIT Press, Cambridge, MA

Nejjari, F. et al (1999). Nonlinear multivariable control of a biological wastewater treatment process. International Journal of Adaptive Control and Signal Processing, Vol.13, 5, pp.347-365

Olsson, G. and Andrews, J. F. (1978). The dissolved oxygen profile-a valuable tool for control of the activated sludge process. Water Research, Vol.12, pp. 985-1004.

Reed, R. D., Marks, R. J. (1999). Neural Smithing: Supervised Learning in Feedforward Artificial Neural Networks. Cambridge, Mass. MIT Press

Rehnström, A. (2000). Automatic Control of an Activated Sludge Process in a Wastewater Treatment Plant- a Benchmark Study. MSc. Thesis, Uppsala University

Rumelhart, E., Hinton.G.E., Williams.R.J. (1986). Learning internal representations by error propagation in Parallel Distributed Processing. Vol.1.MIT Press. Cambridge.pp.318-362.

Samuelsson, P. (2001). Modeling and Control of Activated Sludge Processes with Nitrogen Removal. PhD Thesis, Uppsala University

Seborg, D.E., Thomas F.E., Mellichamp, D.A. (1989). Process Dynamics and Control. John Wiley, New York, N.Y.

Soderstrom, T. and Stoica P. (1989). System Identification. Prentice Hall, Upper Saddle River, N.J.

Tchobanoglous, G., and Burton, F.L. (1991). Wastewater engineering, treatment, disposal and reuse, McGraw-Hill Publishing Company, New York, N.Y.

Vanrolleghem, P. A., Jeppsson, U., Carstensen, J., Carlsson, B. and Olsson, G. (1996). Integration of wastewater treatment plant design and operation - A systematic approach using cost functions. Water Science and Technology, Vol.34 (3-4), pp.159-171.

Vanrolleghem, P.A. (2000). The use of models in advanced WWTP control - Not only model-based predictive control. In: Proceedings TI-KVIV Studiedag Geavanceerde Controletechnieken in de Waterzuivering. Gent, Belgium, Oct. 25, 2000. pp. 25.

Wasserman, D. (1989). Neural Computing: Theory and Practice. Van Nostrand Reinhold, New York.

Watts, J. (1995). Industrial biological wastewater treatment plants, Minworth Systems Ltd, http://www.msluk.demon.co.uk/MSL%20Papers/MSL_Papers_Home.htm Dec. 20, 2002

Zhu, J. et al (1998). An on-line wastewater quality prediction system based on a time-delay neural network. Engineering Applications of Artificial Intelligence, Vol.11, pp. 747-758.

APPENDIX 1

DETAILED DESCRIPTION OF WWT VARIABLES

Refer to Figure 1.1, the detailed illustration and discussion of each component is given below.

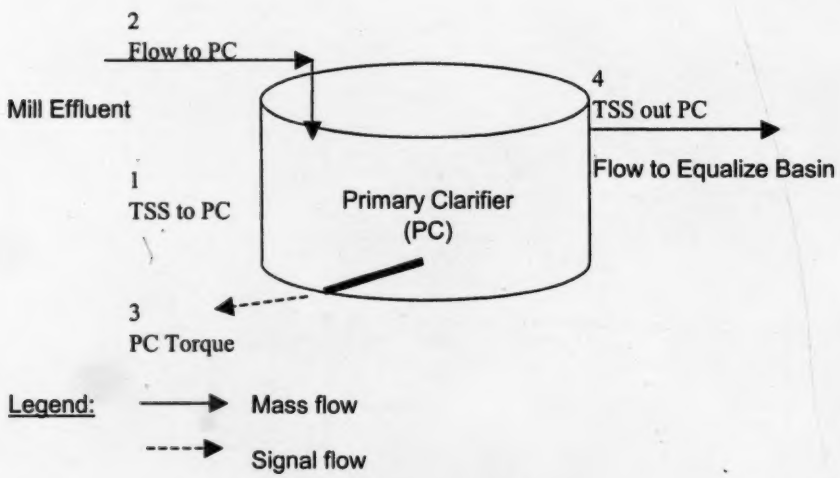


Figure A1.1 Primary Clarifier

As shown in Figure A1.1, the function of the Primary Clarifier (PC) is to remove some of the settleable solids, which cannot be filtered by preliminary treatment. The screened effluent will flow to a top feed. The wastewater is detained in the clarifier for several hours. This enables gravity to act on the suspended particles. Particles heavier than water sink to the bottom of the clarifier and form primary sludge. This sludge is

removed by the collector arm at the bottom and pumped to the solids-handling area of the plant. The primary overflow, also called primary effluent flows to the equalize basin (EB). The variables shown above are illustrated as follows.

TSS to PC, variable #1, measures the TSS before effluent enters PC;

Flow to PC, variable #2, measures the flow rate to PC; It is measured by a magmeter using electromagnetic fields generated across the meter's flow tube.

PC Torque, variable #3, measures the drive force of the sludge skimming equipment needed at the PC. Torque transducer with inductive angular position measurement system measures the torques.

TSS out PC, variable #4, measures the TSS when treated effluent leaving PC;

The TSS is a parameter used to measure water quality as a concentration (weight of solids/volume of water) of mineral and organic sediment. We determined TSS by measuring the weight of dry solid material remaining after vacuum filtration of a known sample volume (50 to 100 mL). Samples were filtered through a 0.45-micron filter in accordance with American Public Health Association protocols.

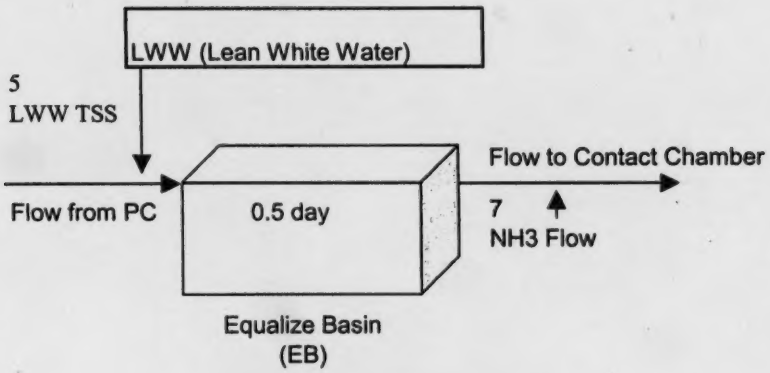


Figure A1.2 Equalize Basin

Fluctuation in mill effluent characteristics can have detrimental effects on the performance of an activated sludge plant. As shown in Figure A1.2, effluent from primary clarifier and lean white water (LWW, process water for treating the pulp, without much organics) enters the basin. With a hydraulic retention of about 0.5 day, the equalize basin allows the basin components to be fully mixed before flowing to the contact chamber. Ammonia is added before the overflow from the equalize basin enters the contact chamber. For the variable shown in Figure 1.2,

LWW TSS, variable#5, is the TSS of lean white water;

NH_3 Flow, variable#7, is ammonia flow rate. The NH_3 is fed to the plant by a metering pump.

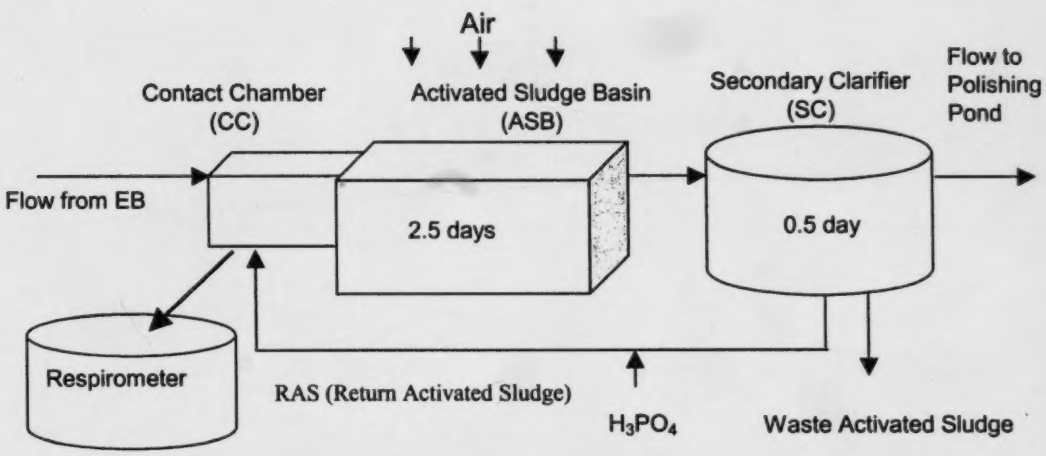


Figure A1.3 Secondary Treatment: ASB, Secondary Clarifier

As shown in Figure A1.3, the function of contact chamber is to allow full mixing of effluent, RAS and nutrients before entering ASB. Phosphorous, in the form of H_3PO_4, is added to the RAS. The RAS is a flow of sludge from the secondary clarifier and is continuously returned to maintain a high population of biomass. The function of each component is described as follows.

Activated sludge basin (ASB): The effluent is retained for about 2.5 days to allow the biomass to grow by using oxygen and organic matter in the wastewater as food. As they grow and are mixed by aeration, the individual organisms stick together (flocculate) to form heavier clumps that thicken;

Secondary clarifier (SC): Effluent from the ASB flows by gravity to the secondary clarifier. The effluent is retained for 0.5 day to allow time for flocculent slurry to form. Part of the settled sludge is pumped back to contact chamber; the rest is taken out of the plant (also called sludge wastage). The level of sludge wastage is adjusted to maintain ASB sludge age and population of biomass at optimum levels;

Respirometer: Taking batch from the ASB, acting as a pilot plant with some similarity to the ASB to monitor the ASB process. For further details, refer to Section 1.4, 5.2;

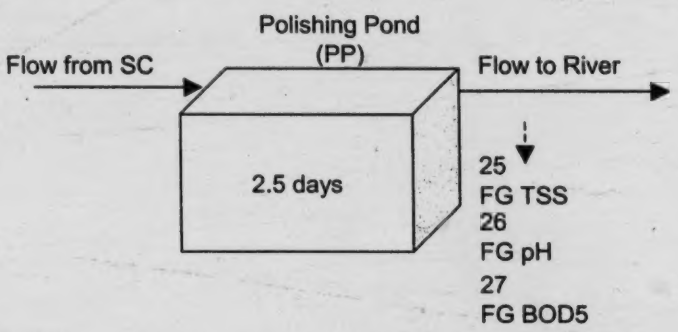


Figure A1.4 Polishing Pond

As shown in Figure A1.4, the function of polishing pond is to allow for some settling of solids carried over from the secondary clarifier or generated in the pond itself before being discharged to river. The effluent is retained for 2.5 days to allow for the settling of solids carried over from the secondary clarifier or generated in the pond before being discharged into the river. For the variable shown above,

FG (Final Grab) TSS, variable#25, is measured at the output of polishing pond before discharged to river;

FG pH, variable#26, is the pH of treated wastewater before discharged to river.

Refer to Figure 1.1, the plots of the WWT variables are presented below. Most of them are sampled once per day. For those variables (all BOD5, NH_3 and P residual) sampled less frequent than daily, the value is given to be zero if there is no data collected on that day.

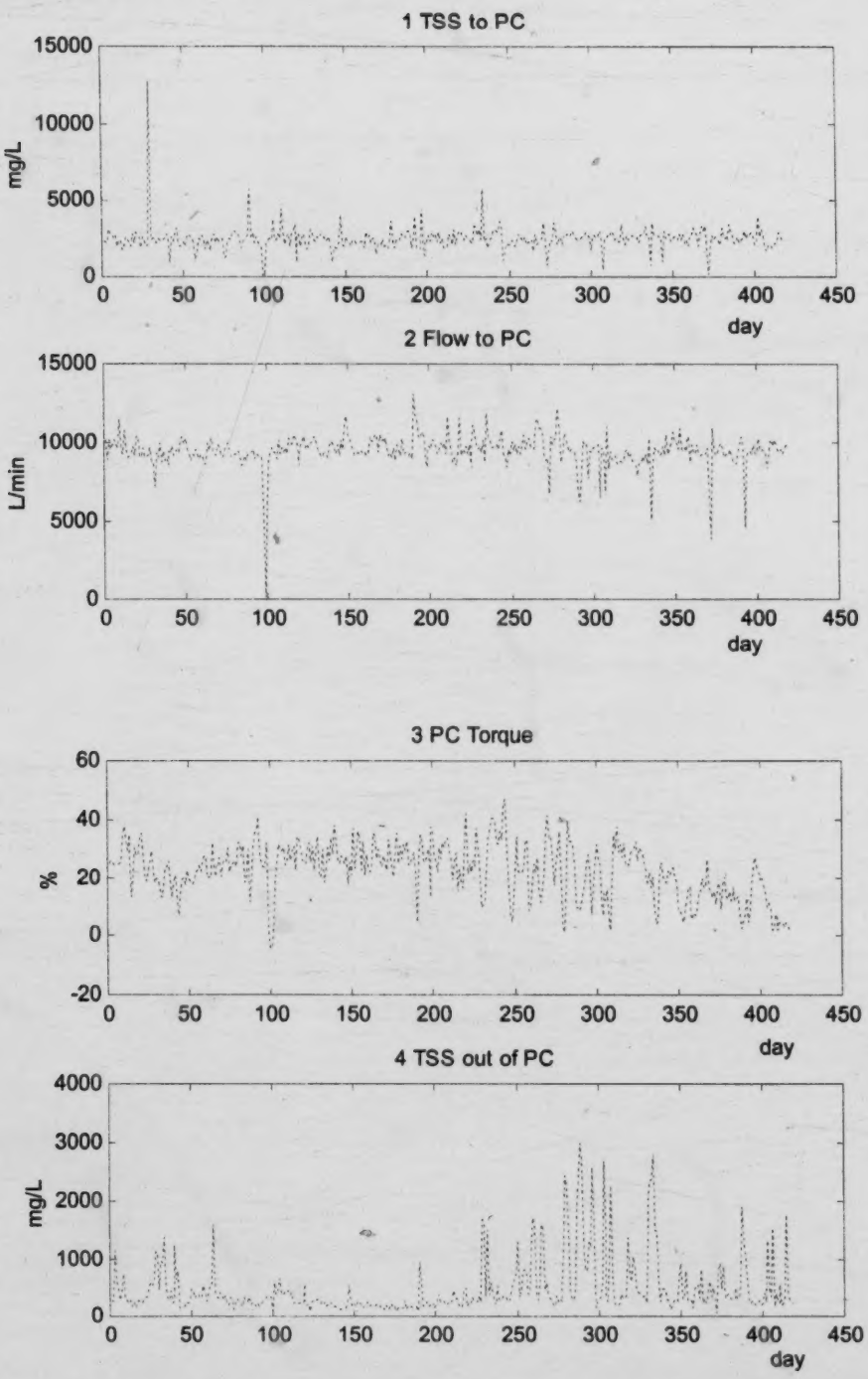


Figure A1.5 TSS to PC, Flow to PC, PC Torque, TSS out of PC

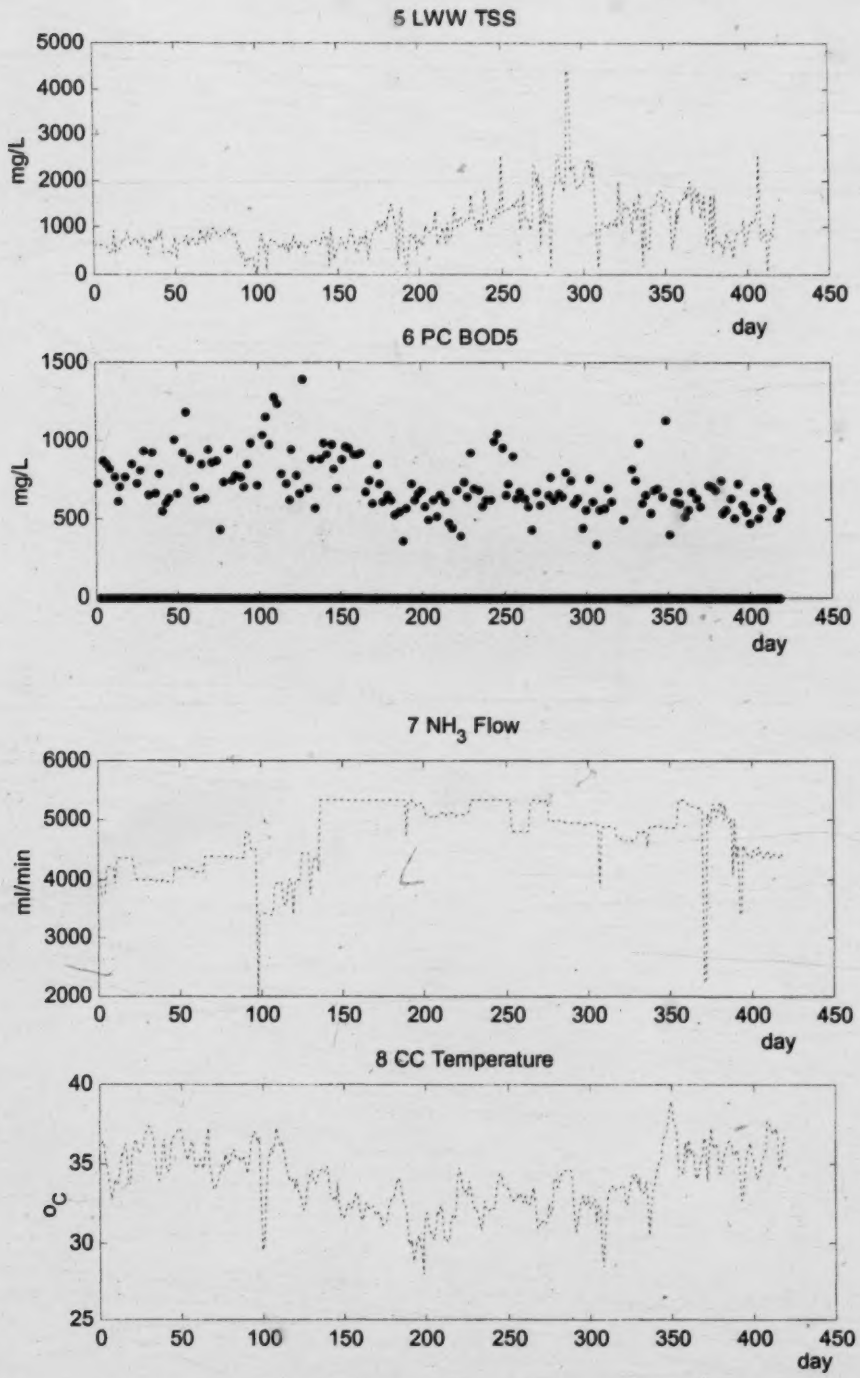


Figure A1.6 LWW TSS, PC BOD5, NH_3 Flow, CC Temperature

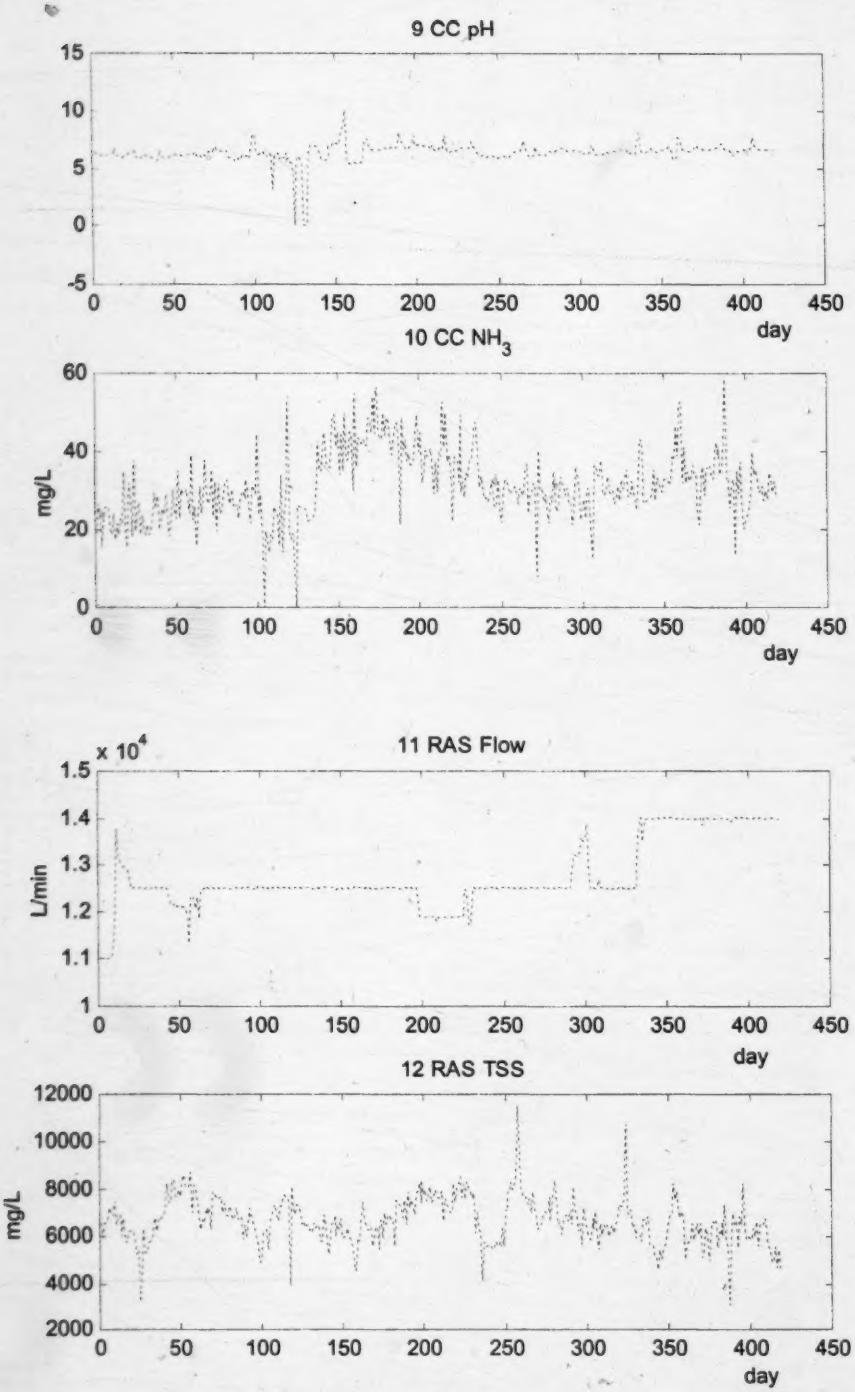


Figure A1.7 CC pH, CC NH_3, RAS Flow, CC NH_3

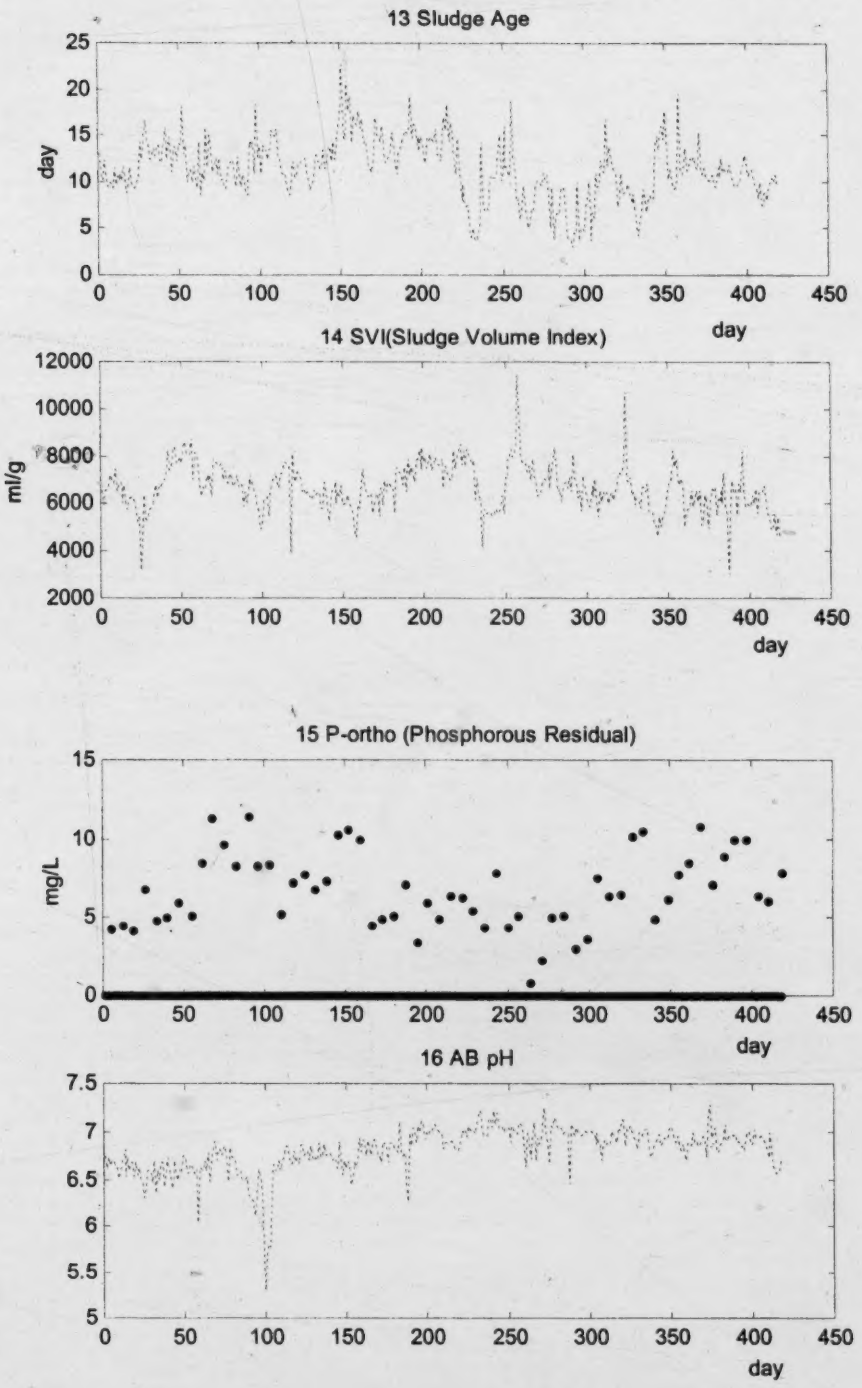


Figure A1.8 Sludge Age, SVI, P-ortho, AB pH

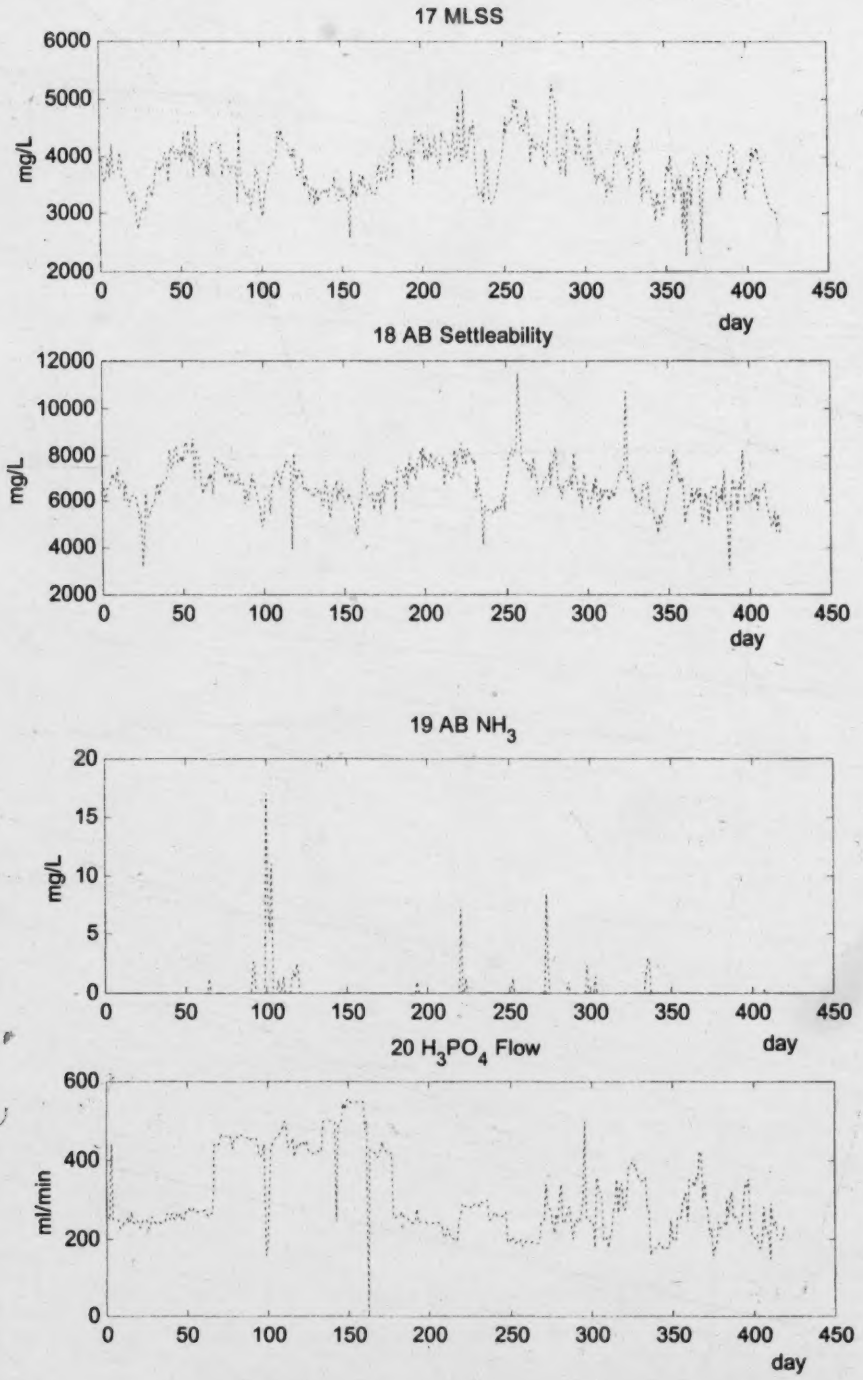


Figure A1.9 MLSS, AB Settleability, AB NH_3, H3PO4 Flow

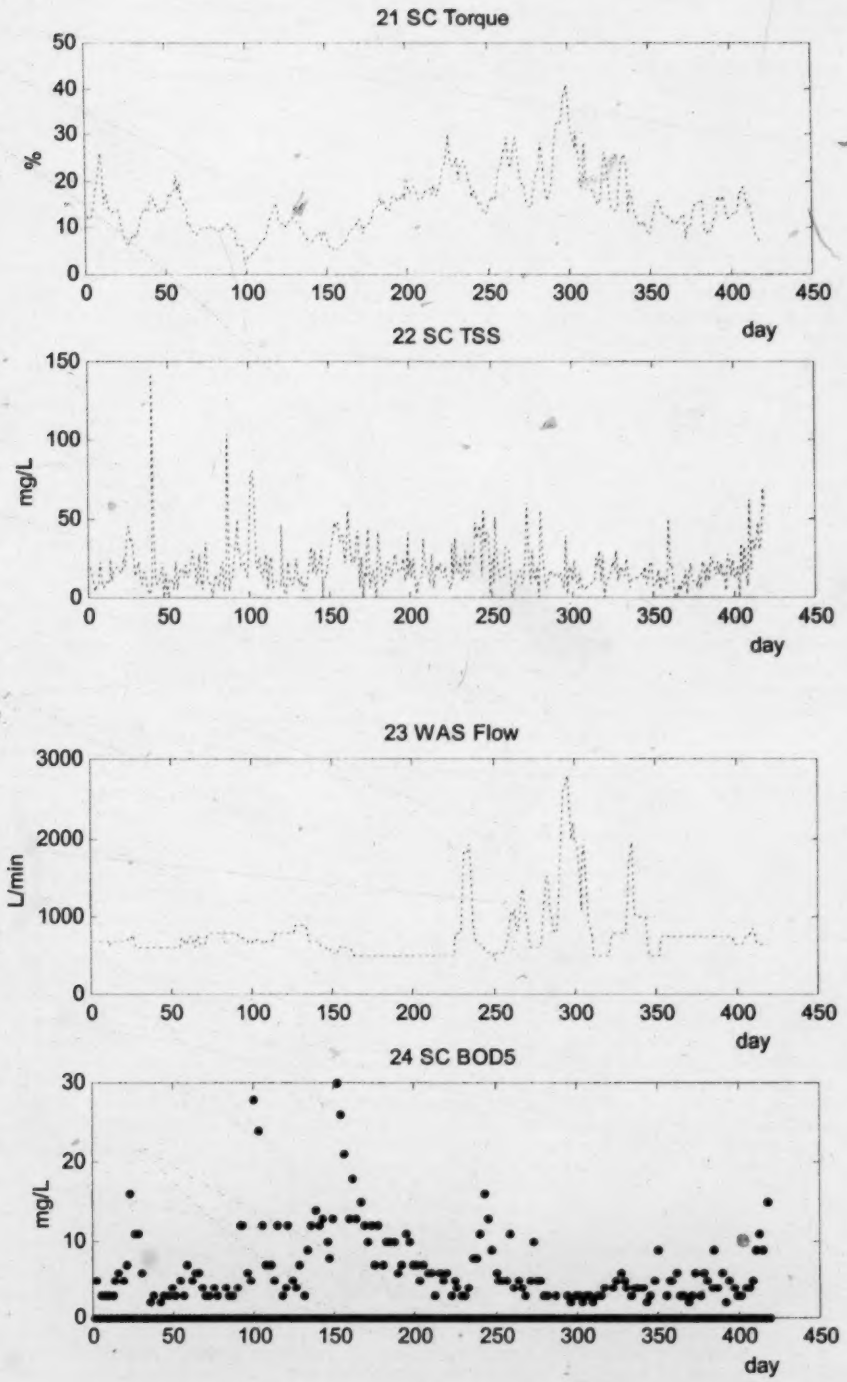


Figure A1.10 SC Torque, SC TSS, WAS Flow, SC BOD5

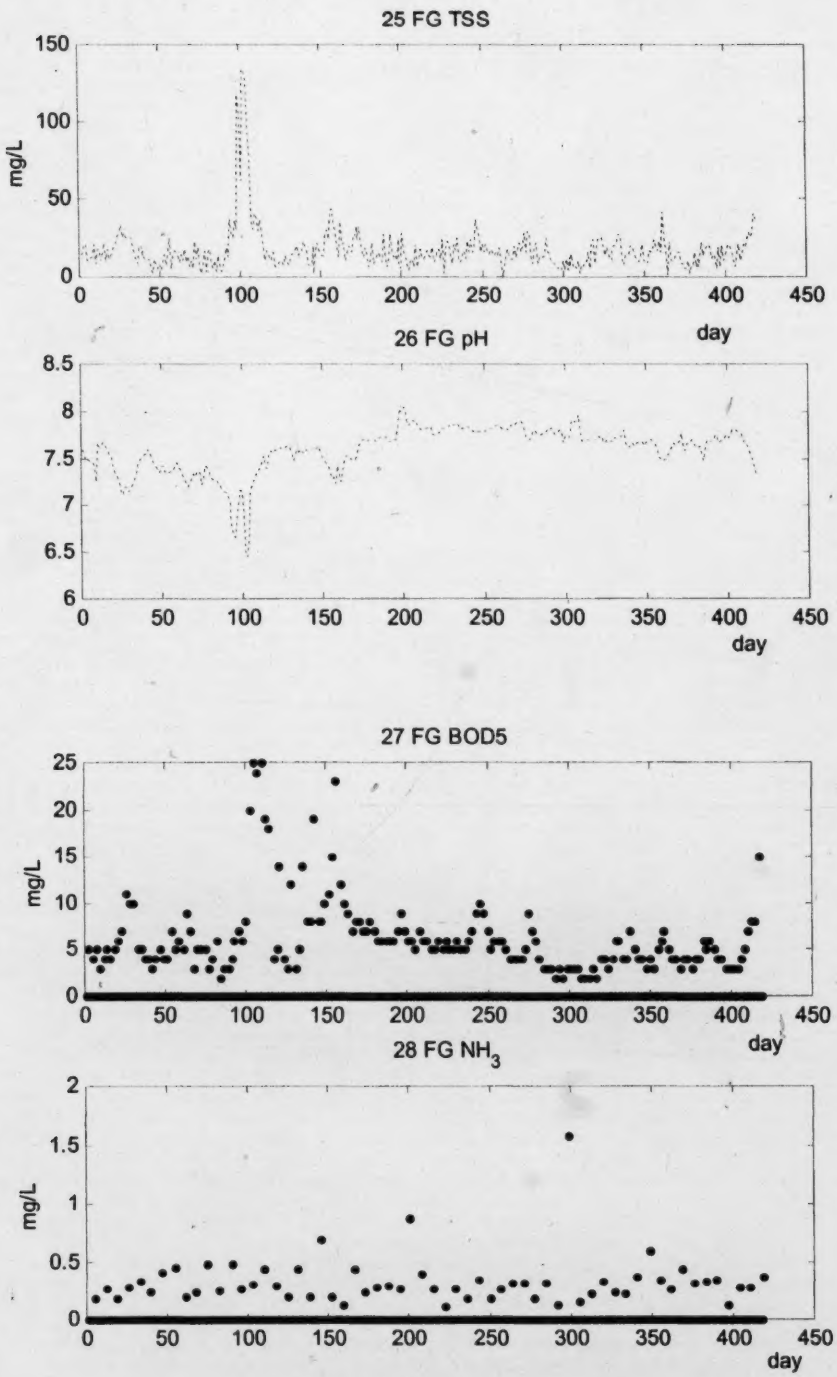


Figure A1.11 FG TSS, FG pH, FG BOD5, FG NH_3

APPENDIX 2

MODELING OF FG TSS BY DIFFERENT DATA PARTITIONING

For the experiments conducted in Chapter 5, investigating the partitioning of the training, validation and testing sets, and separating the data set into training, validation and testing sets representing 50%, 25% and 25%, respectively, of the total data, yielded the following results.

Table A2.1 Performance index / model parameters (one input: AB pH, longer testing set)

Output: FG TSS mean = 15.98		RMSE Training	Validation	Testing	na	nb	nf	nc	nd
		Day 111~ 260	Day 261~342	Day 343 ~419					
One Input									
Input 1:AB pH	ARX	7.72	6.76	7.88	2	7			
	ARMAX	7.76	6.54	7.8	9	7		5	
FG TSS Delay	OE	7.73	6.58	8.05		8	2		
3 days to AB pH	BJ	7.7	6.23	8.06		1	10	4	5

For one input case, Table A2.1 shows similar results as those of Table 5.1. However, in this longer testing set, the model cannot predict a major variance on day of 362.

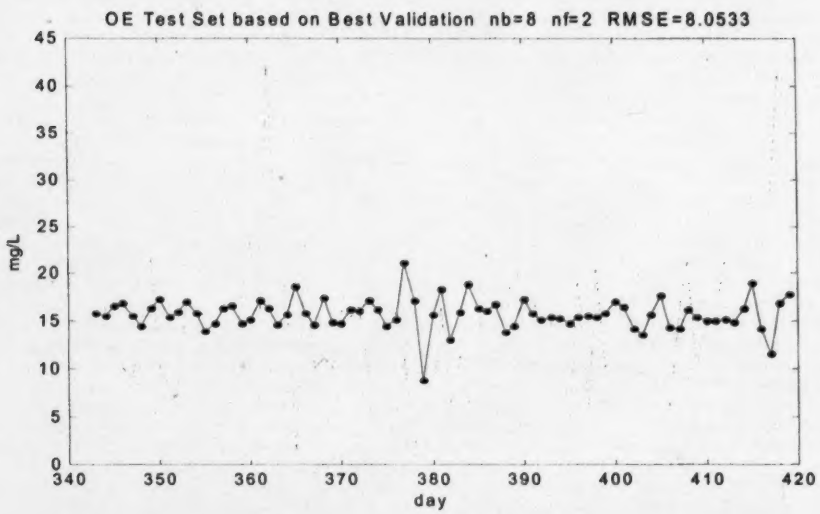


Figure A2.1 OE Longer testing set, one input: AB pH

Table A2.2 Performance index and model parameters (two inputs: AB pH, Settleability, longer testing set)

Output: FG TSS mean = **15.98**		RMSE Training	Validation	Testing	na	nb	nf	nc	nd
		Day 111~ 260	**Day 261~342**	**Day 343 ~419**					
Input 1:AB pH	ARX	7.25	7.12	7.05	4	2			
	ARMAX	7.21	7.11	7.42	5	5		6	
Input 2: Settleability	OE	7.16	7.08	7.42		8	2		
FG TSS Delay 3 days to Setteability	BJ	6.94	7.1	6.92		9	8	9	7
FG TSS Dealy 5 days to Settleability	OE	7.1	7.01	7.42		2	6		

Table A2.2 shows similar results to those in Table 5.2. Comparing Figure A2.2 with Figure A2.1, the two-input model's performance is better than one-input model's.

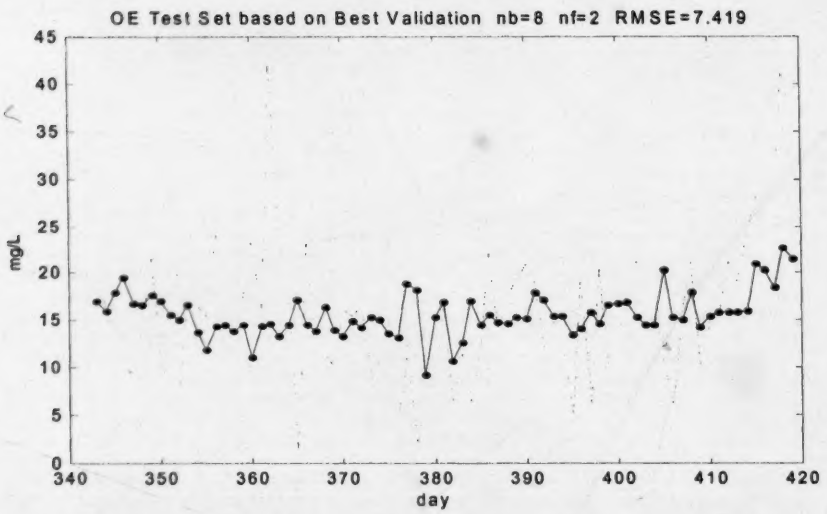


Figure A2.2 OE Longer Testing set, two inputs: AB pH, Settleability

Table A2.3 Performance index /model parameters (three inputs: AB pH, Settleability, OUR shape change, longer testing set)

Output: FG TSS mean = 15.98		RMSE Training	Validation	Testing	na	nb	nf	nc	nd
		Day 111~ 260	Day 261~342	Day 343 ~419					
Input 1:AB pH	ARX	6.96	7.12	7.1	5	2			
	ARMAX	6.75	7.5	7.22	5	6		10	
Input 2: Settleability	OE	6.77	7.32	7.4		6	1		
	BJ	6.66	7.1	8.64		7	7	9	5
Input 3: OUR Shape Change									

For the three input case, comparing Figure A2.3 with Figure 5.21, the OE models' overall performances in predicting the change of FG TSS are similar. The modeling of FG TSS after day 410 is better than that in the two-input case (shown in Figure A2.3). The model can predict the direction of the FG TSS change, especially the change after

day 410. However, the model still cannot accurately predict the change on day 362. The reason may be that the datum on day 362 is abnormal.

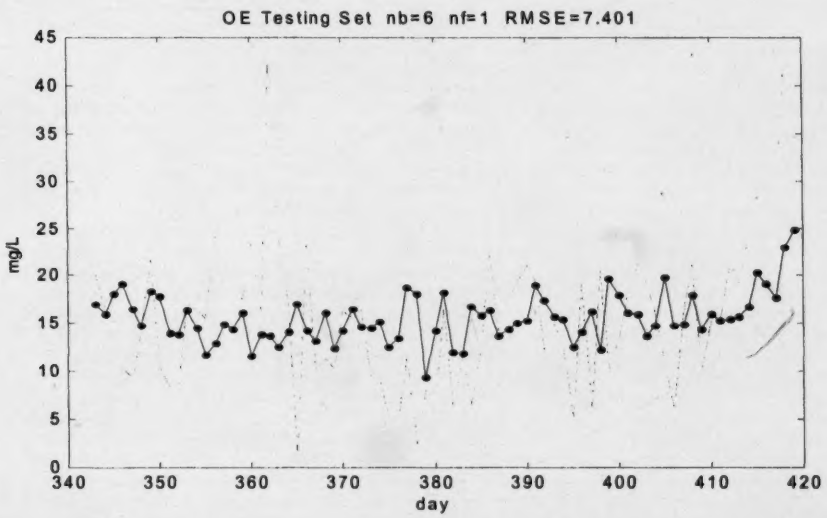


Figure A2.3 OE longer testing set, three inputs: AB pH, Settleability, OUR shape change

APPENDIX 3

USING DATA FROM THE 1ST DAY TO 419TH DAY FOR MODELING

Refer to section 5; rather than using the data from the 108th to 419th day for the linear prediction error modeling of FG TSS, the experiments using the full data set (from the 1st day to 419th day),which includes the spike (happened around the 110th day) are also conducted for all the cases.

For example, the three-input (AB pH, Settleability & OUR Shape change) case's OE model performance index (RMSE) is shown in Table A3.1. Comparing Table A3.1 with Table 5.5, the RMSE of the training set is much higher, but the RMSE of testing set is slightly lower for the OE model.

Table A3.1 Performance index /model parameters (three inputs: AB pH, Settleability, OUR shape change, full data set)

Output: FG TSS mean = 17.48		RMSE Training	Validation	Testing	na	nb	nf	nc	nd
		Day 6~ 218	Day 219~362	Day 363 ~419					
Input 1:AB pH	ARX	19.23	7.69	10.83	5	8			
	ARMAX	15.36	9.49	6.22	6	5		5	
Input 2: Settleability	OE	13.81	12.53	6.25		5	6		
	BJ	17.59	7.43	7.85		5	6	9	6
Input 3: OUR Shape Change									

Furthermore, comparing Figure A3.1 with Figure 5.21, the model output obtained from the training set with the spike can predict the change of FG TSS (after day 410) better than the model obtained without the spike.

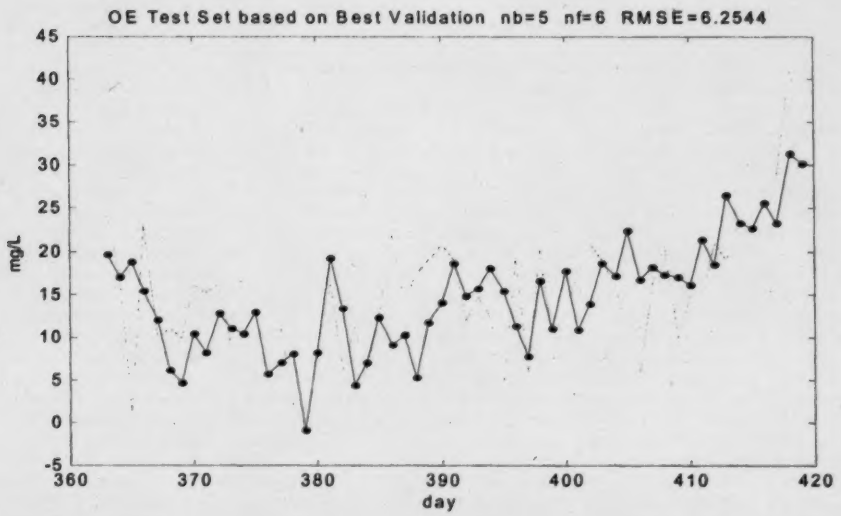


Figure A3.1 OE testing set, three inputs: AB pH, Settleability, OUR shape change, full data set

After taking out the spike, there are only 302 data left (much shorter than 419 data). With less data for training (for the model to learn), the performance of the model is likely reduced. The reason to use the 108^{th} to 419^{th} day data without the spike is to make sure that the data applied in experiment is stationary. From the physical point of view, the spike does not represent a normal operation situation in the whole process, therefore it should be taken out if we're modeling normal operation. Figure A3.2 shows the training set of the OE model with the spike.

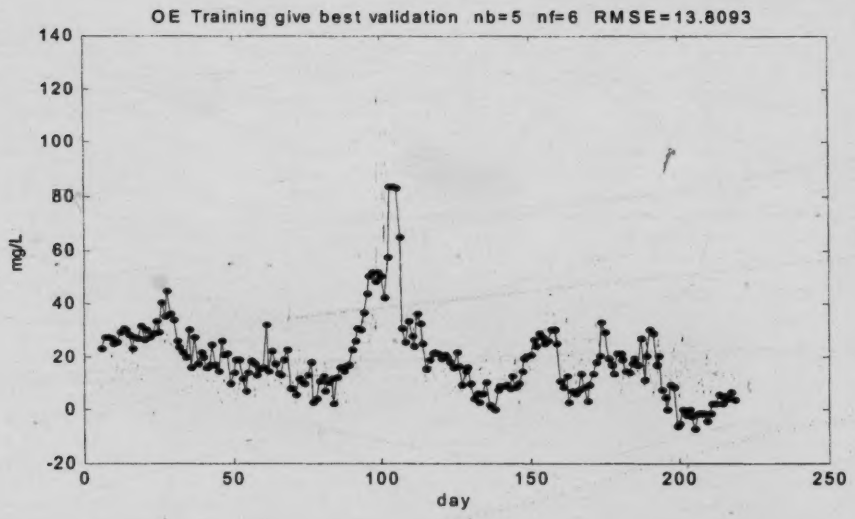


Figure A3.2 The training set of OE model, three inputs, full data set

The detailed description of the application of the above OE model is illustrated below.

The coefficients **B(q)** and **F(q)** of equation (5.1) can be expressed in the matrix form,

$$\mathbf{F} = \begin{bmatrix} 1 & f_1^1 & f_2^1 & f_3^1 & f_4^1 & f_5^1 & f_6^1 \\ 1 & f_1^2 & f_2^2 & f_3^2 & f_4^2 & f_5^2 & f_6^2 \\ 1 & f_1^3 & f_2^3 & f_3^3 & f_4^3 & f_5^3 & f_6^3 \end{bmatrix} =$$

$$\begin{pmatrix} 1.0000 & 1.3110 & 0.9986 & 1.0882 & 0.5446 & -0.0127 & 0.0775 \\ 1.0000 & 0.0298 & 0.0185 & -1.3473 & -0.1436 & -0.0217 & 0.5717 \\ 1.0000 & 1.0058 & -0.2818 & -0.0704 & 0.2368 & -0.2459 & -0.1454 \end{pmatrix}$$

$$\mathbf{B} = \begin{bmatrix} b_1^1 & b_2^1 & b_3^1 & b_4^1 & b_5^1 \\ b_1^2 & b_2^2 & b_3^2 & b_4^2 & b_5^2 \\ b_1^3 & b_2^3 & b_3^3 & b_4^3 & b_5^3 \end{bmatrix} =$$

$$\begin{pmatrix} -19.3412 & -25.6951 & -37.4493 & -46.1984 & -23.0074 \\ -0.0223 & -0.0146 & -0.0118 & 0.0126 & 0.0230 \\ -0.0505 & -0.0815 & -0.0322 & -0.0206 & -0.0183 \end{pmatrix}$$

By inserting the same inputs (inputs start from day 1, shown in Figure 5.25), that is, the AB pH (u_1), Settleability (u_2) value on day 3 and the OUR shape change (u_3) value on day 1, into equation (5.1), we obtain,

$$y(6) = \frac{\mathbf{B(q)}}{\mathbf{F(q)}} \begin{pmatrix} u_1(6-3) \\ u_2(6-3) \\ u_3(6-5) \end{pmatrix} = \frac{\mathbf{B(q)}}{\mathbf{F(q)}} \begin{pmatrix} u_1(3) \\ u_2(3) \\ u_3(1) \end{pmatrix}$$

The operator can use the calculated y(6) to predict the FG TSS on day 6 (shown in Figure A3.3). Accordingly, we can use the input data on day 2 and day 4 to predict the FG TSS on day 7.

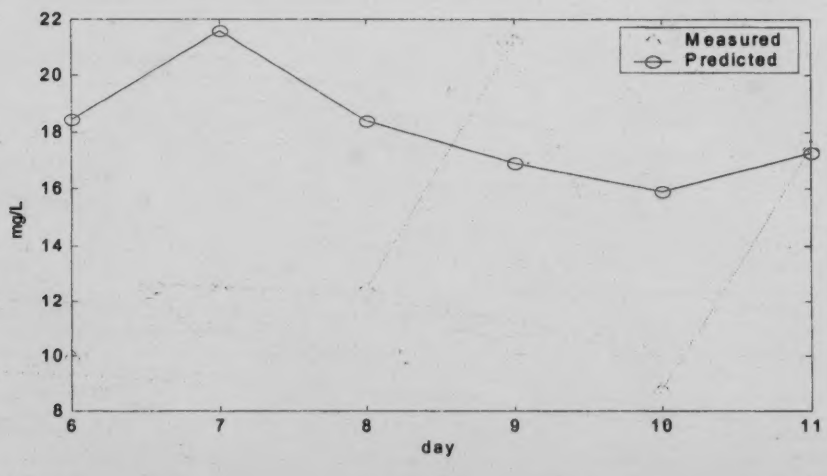


Figure A3.3 Measured and Predicted value of FG TSS, full data set

By different data partitioning, with training, validation and testing are 40%, 35% and 25% of the number of total data, the results are shown in Table A3.2.

Table A3.2 Performance index /model parameters (three inputs: AB pH, Settleability, OUR shape change, longer testing, full data set)

Output: FG TSS mean = 17.48		RMSE Training	Validation	Testing	na	nb	nf	nc	nd
		Day 6~ 205	Day 206~342	Day 343 ~419					
Input 1:AB pH	ARX	19.74	6.95	9.64	6	8			
	ARMAX	14.94	10.64	8.93	5	7		10	
Input 2: Settleability	OE	13.98	12.72	8.22		10	5		
	BJ	18.37	7.54	8.24		5	5	8	5
Input 3: OUR Shape Change									

Although Figure A3.1 shows better performance than Figure 5.21 does, using longer testing set (same length) for both, the model with the spike in the training set gives inferior performance (Figure A3.4) than the model without the spike in the training set (Figure A2.3). It shows that with abnormal data in the data set, the model performance is not stable.

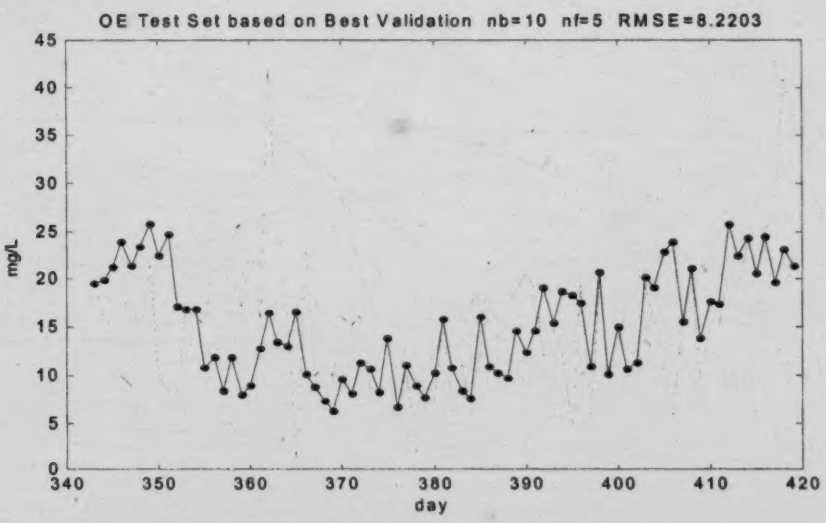


Figure A3.4 OE longer testing set, three inputs, full data set

APPENDIX 4

MULTI-LAYER PERCEPTRON (MLP) MODELING TO PREDICT FG TSS

A 4.1 Theory of Multi-Layer Perceptron Feed-forward Networks model

In many real world processes, we must use incomplete or noisy data and have to make reasonable predictions from the information available. Doing so can be especially difficult without a good theory available to help to reconstruct the missing data. In such situations, back propagation (BP) networks may provide some answers (McAuley, 2003).

A BP network normally consists of three layers: an input layer, at least one intermediate hidden layer, and an output layer (see Figure A4.2). Typically, the nodes are connected in a feed-forward fashion with input nodes fully connected to nodes in the hidden layer and hidden nodes fully connected to nodes in the output layer. When a network is cycled, an input pattern is propagated forward to the output units through the intervening input-to-hidden and hidden-to-output weights (McAuley, 2003).

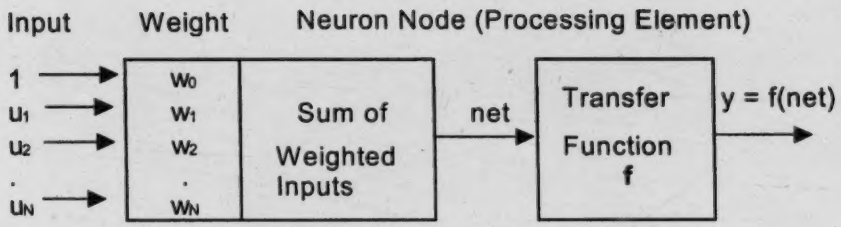


Figure A4.1 Neuron Node structure

Figure A4.1 shows the basic neuron structure with N inputs. Each input is weighted. The sum of the weighted inputs is net= $w_0 + w_1u_{1+},..,+w_Nu_N$ ($w_0,w_1,..,w_N$ are weights form the inputs to the transfer function f). The output of the neuron is y=f(net).

As a classical neural network, an MLP network acts as an approximator through back propagation with structures capable of universal approximation where knowledge and function are distributed among the nodes, each performing some simple (usually non-linear) computation. The network is normally optimized by as follows. First, present a training pattern and propagate it through the network to obtain the outputs; Second, compare the outputs with the desired values and calculate the error; Third, calculate the derivatives of the error with respect to the weights; Fourth, adjust the weights to minimize the error; Finally, repeat until the error is acceptably small or time is exhausted (Reed, 1999). As an application in this thesis, the structure of the MLP for pattern recognition is shown in Figure A4.2. This structure has three inputs and one hidden layer with M hidden nodes.

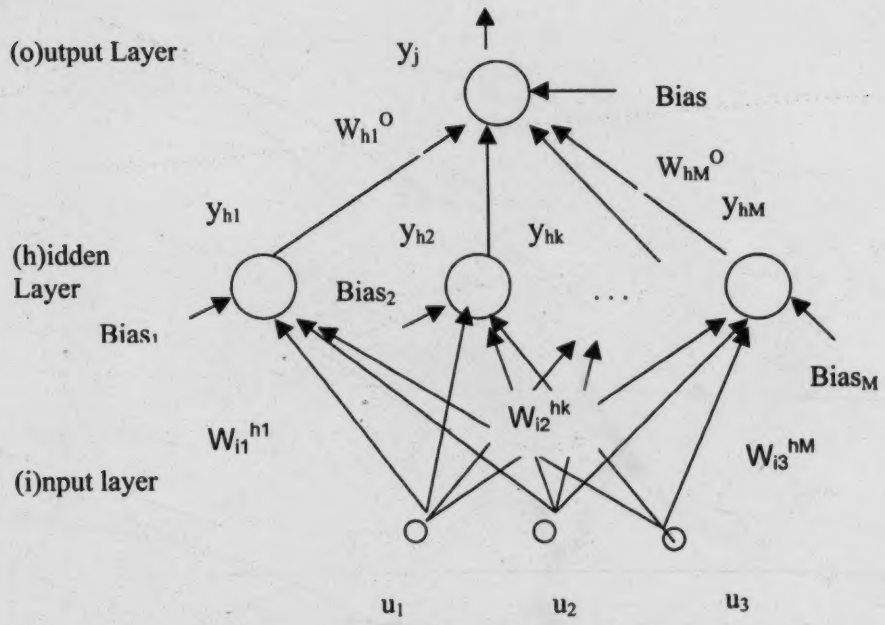


Figure A4.2 MLP approximator structure

Two kinds of nodes are applied in the MLP structure, the hidden node and the output node. For the hidden nodes, the bipolar activation function is used:

$$f(net) = \frac{2}{1+\exp(-\lambda * net)} - 1$$

where λ is the sigmoid function coefficient

In the output node, linear transfer function is used:

$$f(net) = net$$

In the case of one output, y is to be approximated by the approximation error as

$$E = \frac{1}{2}\sum_{j=1}^{J}(t_j - y_j)^2$$

where t_j is thetarget, the real measured value, and

y_j is the output from the output neuron, the prediction by MLP, and

J is the number of input-output points

Since $y_j = f(net^o)$,

where net^o is the sum of weighted inputs to output neuron

$$net^o = Bias + \sum_{k=1}^{M} y_{hk} \cdot w_{hk}^o$$

apply the delta rule,

$$\delta^o = (t_j - y_j) \cdot f'(net^o)$$

From one node (e.g., the k^{th} neuron,) in the hidden layer to the output node, the weight update from the hidden layer to the output layer is

$$\Delta w_{hk}^o = \eta \cdot \delta^o \cdot y_{hk}$$

that is,

$$\Delta w_{hk}^o = \eta \cdot (t_j - y_j) \cdot f'(net^o) \cdot y_{hk}$$

where y_{hk} is the output of the k^{th} hidden neuron

w_{hk}^o is the weights from the first node in the hidden layer to the output, and

η is the learning rate

Therefore, the weights from the hidden to the output node are updated as

$$w^{o}_{hk} = w^{o}_{hk} + \Delta w^{o}_{hk}$$

since

$$y_{hk} = f(net^{hk})$$

while

$$net^{hk} = Bias_k + \sum_{p=1}^{N} u_p . w^{hk}_{ip}$$

where $Bias_k$ is a constant,

N is the number of inputs, N=3 in this case, and

u_p is the p^{th} input data to the neural network

Similarly, the weights are updated from the input layer to the hidden layer,

$$\Delta w^{h}_{i} = \eta \cdot \delta^{h} u_p$$

$$\delta^{h} = f'(net^{h}) \cdot \delta^{o} w^{o}_{h}$$

Therefore, the weights from the input to the hidden node are updated as

$$w^{h}_{i} = w^{h}_{i} + \Delta w^{h}_{i}$$

where w^{h}_{i} is the weight from the input to the hidden layer.

The following example is an application of BP method to model the function $y(x) = 10\sin\left(\frac{2\pi \cdot x^2}{200}\right)$, where x is in a sample range of [0,20] with an interval of 0.5. Optimize the approximator with respect to the RMSE by adjusting the number of hidden nodes. The structure of the network is shown in Figure A4.3.

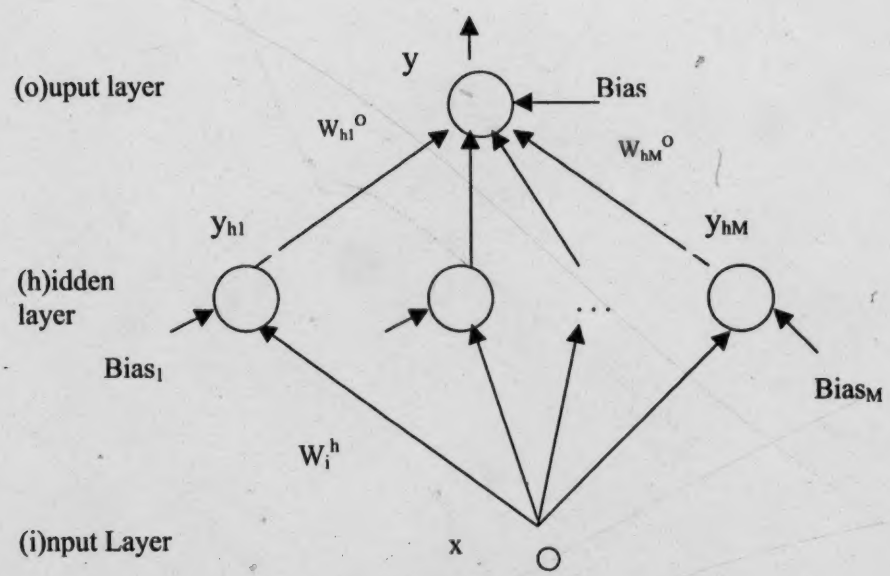


Figure A4.3 MLP structure with one input, one output

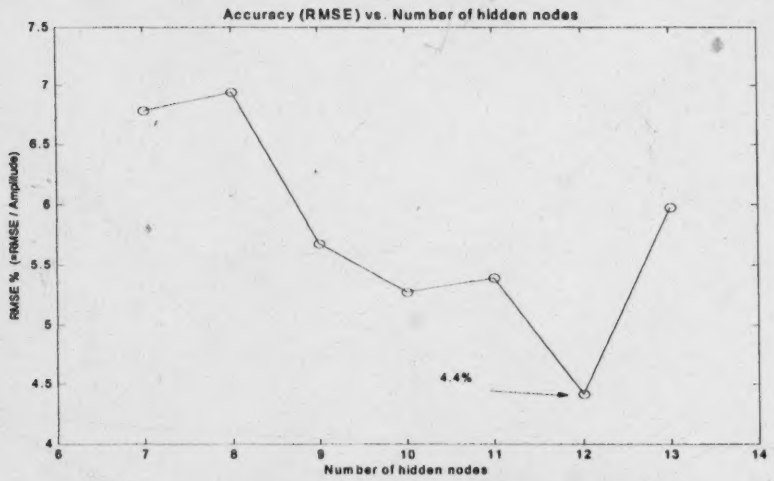


Figure A4.4 RMSE vs. Number of hidden nodes

As shown in Figure A4.4, with 12 hidden neurons and proper initialization of weights, after training of 3000 epochs, the BP method gives an RMSE of **4.4%**(0.44/10).

The approximation and function output of the whole data are shown in Figure A4.5,

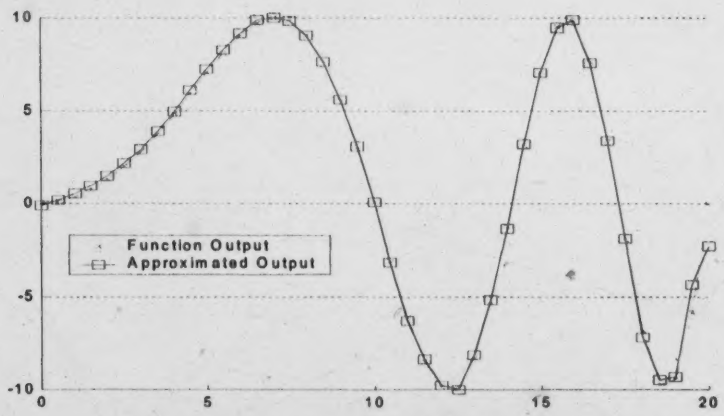


Figure A4.5 Function vs. Model outputs

As shown in Figure A4.5, through BP, the neural network model output indeed fits the given function's output very well.

A 4.2 MLP modeling applied to predict FG TSS

The static MLP modeling is tested as follows. With AB pH, Settleability and OUR shape change as three inputs, the prediction of FG TSS by MLP is shown in Figure A4.6. The training, validation and testing set are consolidated in one plot. All inputs and output are normalized to (0,1) by formula $x = \frac{x - \min}{\max - \min}$ in modeling.

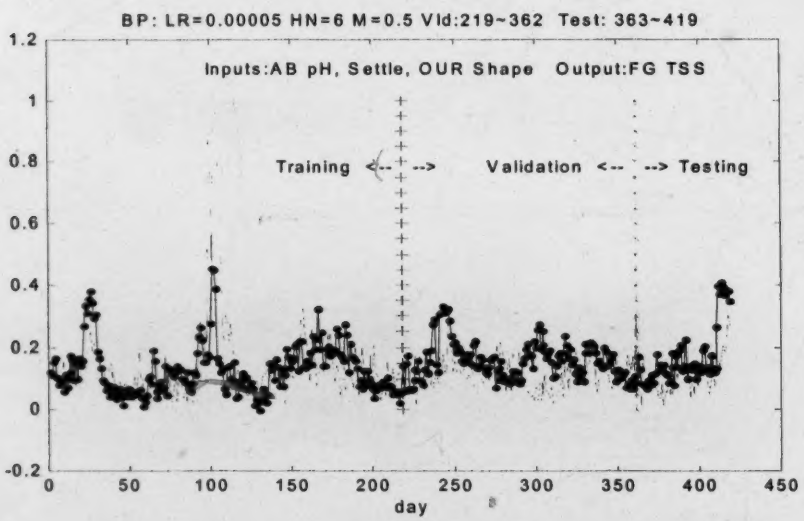


Figure A4.6 Plot of FG TSS and Prediction of all periods

When use RMSE for comparison, the RMSE value of FG TSS prediction is rescaled back by multiply a coefficient of 134.5 (=TSS_max-TSS_min) and add the minimum value of TSS (TSS_min).

The training set is shown in Figure A4.7, and the MLP model fits the FG TSS not so well as the OE model did, which is shown in Figure A3.2.

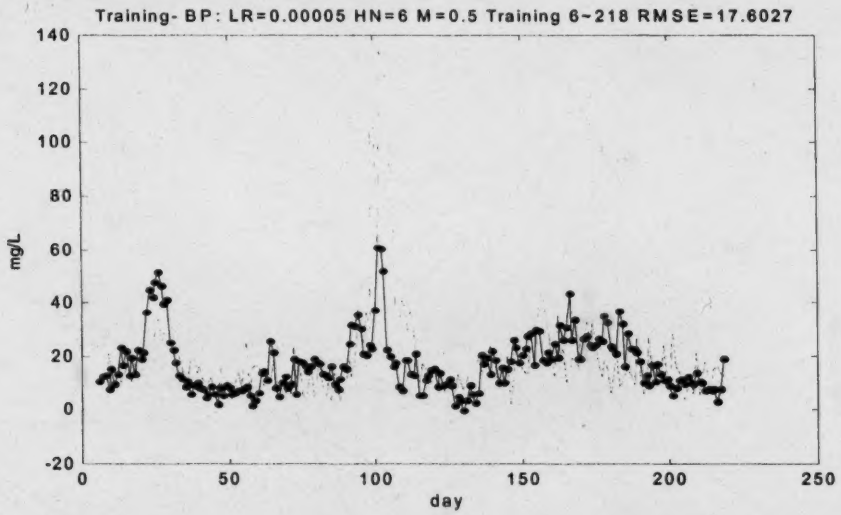


Figure A4.7 MLP Training set, three inputs: AB pH, Settleability, OUR shape change, full data set

As shown in Figure A4.8, the testing set, an independent data set, can verify the model performance without any bias. The MLP model can predict the change of FG TSS to certain extend. The model applied is a static MLP model. The performance of a linear prediction model with RMSE of 6.25 (Figure A3.1), which is a dynamic model, is better than the performance of the static MLP model with RMSE of 12.30 (Figure A4.8).

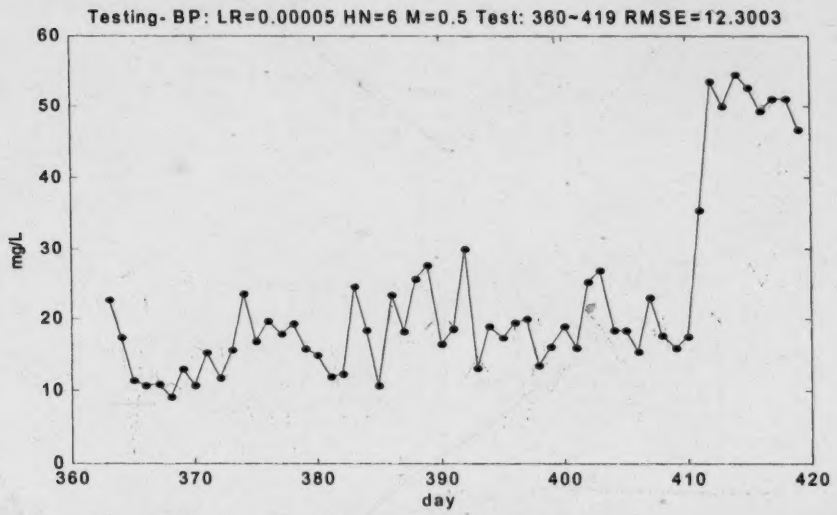


Figure A4.8 MLP Testing set, three inputs: AB pH, Settleability, OUR shape change, full data set